A cross-cultural comparison of the role of maternal mental health in the prediction of infant cognitive development and empirical investigation of the role of early caregiving in India.

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by Matthew Bluett-Duncan.

February 2021

Table of Contents

Total word count (including appendices): 93,177.

List of Contents

List of Figures

List of Tables

Abstract

A cross-cultural comparison of the role of maternal mental health in the prediction of infant cognitive development and empirical investigation of the role of early caregiving in India.

Matthew Bluett-Duncan, Doctor of Philosophy (PhD).

Background The association between postnatal depression and cognitive development has been extensively researched in high-income countries (HICs). The results of a systematic review conducted as part of this thesis indicated that more high-quality longitudinal studies are required to examine the association between postnatal depression and infant cognitive outcomes in LMICs. An empirical investigation was therefore conducted in India (Study 1) to evaluate this relationship and to examine the potential moderating role of early maternal sensitivity. Additionally, cross-cultural comparison of depression symptoms are increasingly common in the literature but differential item functioning (DIF) between groups may result in scale items taking on different characteristics across cultures and lead to skewed findings. This thesis describes the development of a set of anchoring vignettes (AVs) and their application to an Indian and UK dataset so that the impact of DIF on a popular self-report tool assessing perinatal depression could be determined and accounted for (Study 2). **Methods** Study 1 data was drawn from an existing longitudinal, community-based sample from India and the associations between exposure to early postnatal (n = 309) or chronic (n = 395) depression and infant cognitive development were examined. Mothers completed the Edinburgh Postnatal Depression Scale (EPDS) at 8 weeks postpartum and then at 6, 12 and 24 months of age, and infants were assessed using the cognitive and language subscales of the Bayley Scales of Infant Development (BSID-III) at 24 months. Maternal sensitivity was assessed and coded using the NICHD mother-infant interaction scales at 6 months. In the second part of the thesis, a set of anchoring vignettes (AVs) were developed to detect DIF on EPDS items between urban populations in India and the UK. In India, data was drawn from the same longitudinal cohort as study 1. 247 mothers completed the AVs, and 549 mothers completed the EPDS at 12 months. In the UK, 828 mothers from an existing longitudinal, community-based sample completed the EPDS at 12 months, and 252 mothers from a separate community-based sample completed the AVs. **Results** Study 1 revealed a significant small association between early postnatal depression and language subscale scores. Following adjustment for covariates neither postnatal nor chronic maternal depression significantly predicted scores on either the cognitive or language subscales of the BSID-III. There was also no evidence of a moderating effect by maternal sensitivity. However, there was a significant positive association between maternal sensitivity and language scores for boys only. In Study 2 the AV analyses indicated that UK mothers rated the vignettes and themselves more severely than Indian mothers. Prior to DIF adjustment, rates of depression were significantly higher in UK participants, but after DIF adjustment, rates were significantly higher in Indian participants. **Conclusions** Study 1 indicates that postnatal depression may not have the same impact on infant cognitive development in India as seen in HICs. However, in line with research in HICs, there was evidence that boys may be more reliant than girls on the early regulation provided through sensitive maternal caregiving. The findings also highlight the need to investigate the potentially protective or adverse roles of other nutritional or socio-cultural factors specific to this context. The results from Study 2 indicate that UK mothers have higher expectations regarding postnatal depression than Indian mothers and so give higher ratings to their own symptoms. The findings indicate that the EPDS does not function equivalently in these settings and extends the literature in support of AVs as a valid and helpful data harmonisation tool for multi-item scales of mental health. Further, the results suggest that cross-cultural comparisons without AV correction need to be interpreted with caution.

Declaration

This thesis has been submitted to the University of Liverpool and the National Institute of Mental Health and Neurosciences in accordance with regulations for the Dual-PhD agreement between the two institutions. With this exception, no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Dedication

I would like to dedicate this thesis to my family: to Janice, to my Mum and Dad, to Becca and Ben, and to Pete, Sam and Ella. I am so grateful for each one of you.

Acknowledgements

Firstly, I would like to thank my supervisory teams, both in India and the UK, for the immense support and guidance they have provided over the past 3 years. Thank you to Prof Helen Sharp for first taking me on as an MSc student many years ago. Your supervision in various contexts over the years has helped to shape me as a researcher and I will be forever grateful for the opportunities provided to me that have resulted in such personal and professional growth. Thank you to Prof Andrew Pickles, without whom this PhD would have been quite impossible. I have learned much more about statistics in the past year than I thought possible, and I am deeply grateful for your patience in answering all my questions and the time you have devoted to developing and fine tuning the anchoring vignette models. I also consider myself immensely privileged to work with Dr Thomas Kishore and Dr Veena Satyanarayana at NIMHANS. Thank you, Thomas, for looking after me and Janice during our stay in Bangalore. Your guidance in navigating what was a completely new academic system was invaluable, but beyond this, you provided a secure place for me to voice my concerns and opened your home up to us. I am also grateful for the clinical experience I gained under your supervision. Thank you also to Veena for your support whilst at NIMHANS and your invaluable input into my work. Finally, I would like to extend special thanks to Dr Prabha Chandra. Thank you for giving me the opportunity to complete my PhD alongside the BCHADS team. My stay at NIMHANS was a real joy and I will always look back on it fondly. Thank you all for making this experience a possibility.

Thank you to all my friends and colleagues at WCHADS and BCHADS. I have been privileged to work with so many amazing people in the UK and India and I am so grateful for your friendship and the contributions you have all made to the larger studies from which I have drawn a lot of my data. I would like to give special thanks to Dr Nicky Wright who has provided invaluable statistical support for my analysis and answered endless random queries. You are an excellent teacher and I've learned so much from you. Thank you also to Prof Jonathan Hill. I am grateful for your support with the anchoring vignettes and for the opportunities you have provided for me as a PI of WCHADS. Supraja and Laura, I've really enjoyed working with you both as part of the BCHADS team. I'm so grateful for your support over the years and for the friendships we have built through many hours of working together. Thank you to Kay for all those early morning chats over a cup of tea, to Miriam, Stu and Karen for being such pillars of support within the WCHADS team, and to all those others past and present who have contributed to WCHADS. Thank you to everyone at

NIMHANS who supported me, contributed to the translation of the vignettes and went out and collected the data.

I would also like to thank my incredible family and friends. The biggest thank you goes to my amazing wife, Janice. Thank you for all your love, support and encouragement over the past 3 years. Thank you for agreeing to go on this Indian adventure and stepping into the unknown with me. I could not have done this without you and I am constantly grateful that you are in my life. To my parents and to Becca and Ben, I am so grateful to have such a wonderful family. Thank you for all your love and support over the years. I'm looking forward to having a bit more time to see you all now! Thank you to all our other friends and family in the UK who supported us from afar and kept in touch throughout, and to all the friends from around the world that made our stay in India so rich. Finally, thank you to God for your unending faithfulness and kindness toward me, and for your grace which has transformed my life and carried me through every challenge that I have faced.

Preface – Setting the thesis in context.

0.1 Mental health and provision in India

A recent report by the India State-Level Disease Burden Initiative found that, in 2017, 1 in 7 people in India suffered from a mental disorder ranging from mild to severe (Sagar et al., 2020). This same report found that the burden of mental disorders had almost doubled from 2.5% in 1990 to 4.7% in 2017, and the highest disease burden came from anxiety and depression. This marked increase in prevalence may be the result of rapid development in recent decades that has contributed to rising levels of social inequality, rapid changes in social structure and community cohesion, and increased urbanisation, resulting in increasing exposure to associated environmental and psychological stressors (Milner, 2016).

In response to the growing awareness of the burden of mental health, India has begun to take steps to address the apparent need. The Indian government launched the first National Mental Health Policy in 2014 which aimed to promote mental health, prevent mental illness, and eradicate stigma (Sharma, 2014). This was followed up with the revised Mental Healthcare Act in 2017, with the stated objectives of providing *"equitable, affordable, and universal access to mental health care"* (Sagar et al., 2020, p.149). As with all healthcare in India, the implementation of this Act is primarily the responsibility of individual states. There are some advantages to this structure in that it supports a tailored approach that is suitable to the needs of the broad socio-cultural and demographic diversity between states, but it also means there is wide variation in its provision across the country and a key challenge remains in how the intentions of the Act are worked out equally to all Indian citizens (Genjekar, Thekkethayyil & Chandra, 2020).

In terms of maternal mental health, Baron et al. (2016) suggested that this issue has been neglected in many low- and middle-income countries (LMICs), including India. In their study of five different LMICs, they reported that the district of Sehore, Madhya Pradesh, had no mental health services dedicated to perinatal women and had no strategies in place in primary health facilities to detect maternal mental health disorders. This claim is echoed by Uphadhyay et al. (2017) who report that, despite recent developments, maternal mental health is still not a prominent part of India's national mental health programme and that dedicated maternal mental health services are largely deficient in healthcare facilities. More recently, Ganjekar et al. (2020) have argued that India is yet to see the evolution of a comprehensive public health programme for perinatal mental health and there are very

few dedicated perinatal psychiatry services, although there has been a recent increase in calls for the prioritisation of perinatal mental health from professionals within the Indian healthcare system. The authors state that there is a need to raise the level of awareness regarding perinatal mental health amongst obstetricians and other healthcare professionals, alongside the development or adaptation of simple tools for the detection of mental health, as well as the provision of perinatal mental health training to nurses, midwives and doctors.

The complex pathways and potential barriers to accessing available provision present a major challenge to mental healthcare in India (Bagadia & Chandra, 2017). One such barrier is the typical absence of a structured referral system, meaning that the first point of contact for many individuals with mental health disorders may be more likely to be a traditional faith healer or allopathic general doctor than a psychiatrist (Pradhan et al., 2001). Encouragingly, there are indications that this balance is beginning to shift, with a more recent study finding that psychiatrists were more common than faith healers or general medical practitioners as the first level of care for individuals with severe postpartum mental illness (Thippeswamy, Desai, & Chandra, 2018).

0.2 Mental health provision at NIMHANS

The National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore is a multidisciplinary institute that focuses on patient care, postgraduate training, and research. It is one of the premier medical institutes in India and was awarded the status of "Institute of National Importance" by an act of parliament in 2012. As a government institute, it is able to offer care to some of the most vulnerable in society. For example, the 2017-2018 NIMHANS annual report reported that specialised care was provided to 528,552 patients, about 75% of which received treatment at no cost or highly subsidised costs.

As an institute dedicated to pioneering systemic change in the delivery of mental health care, NIMHANS is at the forefront of the drive to improve perinatal mental health provision in India and was therefore an ideal context for this research to be carried out. For example, in 2009, the very first mother-baby unit (MBU) in India was established by the Department of Psychiatry. MBUs, which offer specialised joint care for women who need inpatient care within 12 months of childbirth, are becoming increasingly popular across the world and are now recommended as standard in the UK (National Institute for Health and Care Excellence, 2007). The MBU at NIMHANS is a 5-bed facility that is managed by a team of psychiatrists, clinical psychologists, psychiatric social workers, and nurses, and aims to provide excellent

mental health care for the mother, while also teaching parenting skills, encouraging mother-infant bonding and supporting her spouse and wider family (Chandra, Desai, Reddy, Thippeswamy, & Saraf, 2015). An outpatient perinatal clinic, run by a multidisciplinary team, also offers care regarding mental health issues in pregnancy and postpartum, family and marital counselling, and assessments and interventions for mother-infant bonding. Finally, the NIMHANS Centre for Wellbeing offers clinical services in a community setting for women with pregnancy and childbirth related mental health problems who have been referred by surrounding antenatal and maternal health centres.

In this respect, it is clear to see that progress is being made toward establishing the provision of maternal mental health care. However, the challenge remains as to how these isolated examples are extended to the rest of the country. Ganjekar et al. (2020) stress the importance of promoting research that will generate knowledge to strengthen maternal mental health systems and improve maternal and child outcomes. They argue that the effective use of available resources is essential and outline a number of key strategies for doing this, including developing and validating simple screening methods to identify maternal distress, and prioritising discovery studies to investigate the impact of maternal mental illness on pregnancy and maternal and child outcomes.

0.3 The current study

The current study is split into two main sections. The first section aims to extend what is known regarding the prospective relationship between maternal depression during infancy and infant cognitive development in India. Chapter 1 develops a theoretical framework and provides a detailed review of what is known regarding postnatal depression and infant cognitive development from HIC research. Chapter 2 extends this discussion by providing the first systematic review to explore the relationship between postnatal depression and infant cognitive development. This review also explores the effect of antenatal depression due to the significant links observed between experiences of depression across the whole perinatal period. Finally, chapter 3 investigates the prospective association between early postnatal and chronic depression throughout infancy on infant cognitive development at 2 years in India. As part of this investigation the roles of maternal caregiving and infant sex are also examined. The second section is then focused on improving the methodology used in India in order to facilitate a better understanding of how the Edinburgh Postnatal Depression Scale (Cox, Holden, & Sagovsky, 1987), a common screening tool, functions in this context. This will provide a better understanding of how mothers report maternal

depression in India and to allow for more accurate and valid comparisons to be made with other cultures. This will be done using advanced and novel data harmonisation techniques, thereby providing valuable insight as to the efficacy and applicability of these approaches in this context. Chapter 4 introduces the overall concept of measurement bias and describes the anchoring vignette approach. Chapter 5 describes the development and piloting of a new set of anchoring vignettes for use with the EPDS in the current context and discusses the challenges encountered. In chapter 6, a detailed comparison of two distinct modelling approaches to the anchoring vignettes is provided and the implications of findings regarding differences in response style between the UK and India are discussed. Overall, this body of work aimed at improving the current knowledge and understanding of the role of maternal depression in infant cognitive development in India by providing a specific contextual understanding of the functioning of the EPDS that will allow us to accurately ascertain where these findings and future comparisons sit within the global research context.

0.4 Role of the researcher

0.4.1 Dual PhD – pattern of study

This project was completed as part of the Dual-PhD programme run jointly by the University of Liverpool, UK (UoL) and the National Institute of Mental Health and Neurosciences, India (NIMHANS). The Dual PhD is a 3-year programme with a set pattern of study depending on the candidate's home institute. In this case, I spent the first 2 years of the PhD at NIMHANS and the 3rd year at UoL. While based at NIMHANS, I was part of the Department of Clinical Psychology and placed with the Bangalore Child Health and Development Study (BCHADS) on a day to day basis. As there is a much stronger clinical emphasis at NIMHANS, I completed a number of clinical placements in relevant departments under the supervision of Dr M. Thomas Kishore alongside his research work. I also participated in a broad array of departmental programs, including seminars, clinical case presentations and journal clubs. Upon returning to UoL, I was part of the Institute of Health and Life Sciences and placed with the Wirral Child Health and Development Study (WCHADS) supervised by Professor Helen Sharp. Prior to taking up my Dual PhD Scholarship my research career had begun as a research assistant on WCHADS and the design of the BCHADS was based on that study with close collaboration between the study investigators to ensure planned parallel measurement from pregnancy to age 2.

0.4.2 Development of research question

Prior to joining the dual-PhD programme, I held initial discussions with the PIs of the BCHAD study regarding the broad focus of the research question in order to develop a mutually beneficial project. A proposal to focus on cross-cultural measurement calibration via the anchoring vignette methodology was agreed upon and formed the basis for the full research protocol which was developed while at NIMHANS.

As this PhD involves secondary data analysis attached to an already established longitudinal study (BCHADS), there were certain constraints on the focus of the project. The detailed focus of the protocol was therefore primarily developed as I familiarised myself with the BCHAD study after arriving at NIMHANS, in the context of continued discussion with my supervisory team in both institutions and the study investigators.

0.4.3 Anchoring vignette development, data collection and analysis

My primary role in data collection was the development of the anchoring vignettes for the Edinburgh Postnatal Depression Scale (EPDS). Anchoring vignettes are a novel methodology that have not been used in this context or with this measure previously. As anchoring vignettes have not been widely used, there are no ready-made guides to developing them. The first task, then, was to develop an understanding of how the methodology worked and to draw together all the relevant literature to produce a systematic process of vignette development and translation. Following this, a large proportion of my time was spent developing the actual vignettes and then co-ordinating an international translation process to ensure that the vignettes were translated according to World Health Organisation guidelines. Once the anchoring vignettes were finalised, I oversaw their administration to participants by the BCHADS research team. This included programming a version of the questionnaire that could be administrated on offline tablets in the community, training the research team, and continually tracking and reviewing data collection progress. On my return to the UK a comparison sample of anchoring vignette data was collected in the UK. I developed this as a separate study protocol in conjunction with the UK supervisory team and submitted to the UoL ethics committee. Data collection was completed online and recruitment to the study was overseen by myself. Due to the complexity of the anchoring vignette analysis, it was agreed that the expertise of Professor Andrew Pickles (co-supervisor) would be needed to develop the model scripts for use in Stata. In this respect, my role in the anchoring vignette analysis was the preparation of the data and the interpretation of results. In this respect, it was a requirement that I was able to understand and discuss the theoretical basis of each model.

All other data used in this study was collected as part of WCHADS or BCHADS by the relevant research team. Alongside my own work, I have also contributed to BCHADS and WCHADS by developing anchoring vignettes for two other measures (CBCL/1.5-5, STAI) and various data entry, cleaning and analysis tasks. I have also been involved in preparing for the current wave of BCHADS assessments and have been working with the team to develop an online version of the overall questionnaire pack that can be used for maternal assessments.

Chapter 1: Background

1.1: Cognitive development during infancy

1.1.1: Cognitive development

Cognition can be defined as the mental processes by which the knowledge required for everyday living is acquired, retained and used (Magni & Bilotta, 2016). The word 'cognitive' comes from the Latin word 'cognoscere', which means 'to know'. It is an umbrella term that encompasses the whole range of mental activities and skills, including memory, language, learning, problem solving, perception, concepts, metacognition, and social cognition (Cognitive Development, 2020; Kihlstrom, 2018). The study of cognitive development focuses on how these internal mental processes grow and change throughout the lifespan, particularly with regards to how information is acquired, processed, and organised (Oakley, 2004; Taylor, 2005). During infancy cognition is typically assessed in terms of developmental milestones, including motor actions and language skills, and is conceptualised broadly as cognitive development, rather than the specific higher order functions, such as executive function and complex problem solving, that emerge later (Guerra, Williamson, & Lucas-Molina, 2012). In order to remain consistent and facilitate comparison with the existing literature, this project utilises an inclusive and focused conceptualisation of global infant cognitive development. Additionally, due to the hypothesised interdependent and inter-related nature of early cognitive and language milestones (Deak, 2014), both domains will be considered, reviewed and examined under the label of 'cognitive development'.

1.1.2: Long-term stability of cognition

Population studies provide considerable evidence of the continuity of cognitive development and intelligence from infancy to later life. Blaga et al. (2008) found clear and strong evidence of continuity of intellectual development within and between instruments from 1 to 4 years. Girault et al. (2018) found that scores on cognitive assessments at age 2 were significant predictors of intelligence scores at age 6 in a large heterogeneous sample. Results from Rose et al. (2011, 2012) provide evidence of continuity of core cognitive abilities from infancy (7 and 12months), through toddlerhood (24 and 36 months), to preadolescence (11 years). Finally, Fagan, Holland, and Wheeler (2007) demonstrated that an infant's information processing ability at 7-12 months predicted IQ and academic achievement at 21 years. However, while cognition does appear to be relatively stable throughout the lifespan, there is also evidence that development is susceptible to changes in the environmental context, particularly throughout pregnancy and early infancy.

1.1.3: Infancy and early childhood as a sensitive period

Early childhood represents a key period for the developing brain and is thus of great interest to those studying cognitive development. The perinatal period, and beyond into infancy, has received particular attention because it represents an important and sensitive phase where children are at their most receptive stage of development (Black et al., 2017; Daelmans et al., 2017; Doyle, Harmon, Heckman, & Tremblay, 2009). From conception, up until around 2 years, the brain exhibits a high level of neural plasticity and rapid synapse formation that enhances the capacity of the child to learn and develop (Hannon, 2003). During this period neural connections are formed that affect development throughout the life course and provide the basis for future social, emotional and cognitive development (Sameroff, 2010). While this neural plasticity is beneficial for development under optimal conditions, adverse life experiences during this time can have long-lasting, detrimental effects (Teicher & Samson, 2016). As infancy represents a period of increased vulnerability, early adverse experiences can result in *"long-term physiological and epigenetic effects on brain development and cognition"* (Black et al., 2017, p.77; Chen & Baram, 2016; Daelmans et al., 2017).

The sensitive period hypothesis is difficult to test because normal samples may lack the variability to detect an effect and, in high-risk samples, risk often persists beyond the period in question and so can confound results. A set of studies which avoid these methodological pitfalls are now discussed. The first focus on the child's early experience of having a cleft-lip and the second focus on a ground-breaking set of studies involving children in institutionalised care. Both of these represent singular risks that are ameliorated at a specific point in some children but not others, thereby providing sufficient variation in early adverse experience.

Using a sample of healthy control children and children with cleft-lip who had surgery at 3-4 months, Murray et al. (2008) found that differences in the quality of mother-infant interactions between the groups at 2 months accounted for differences in cognitive development at 18 months. Following surgery, there were no differences in interaction quality between the two groups, meaning the association with cognitive outcome was specific to the pre-surgery interactions. When children's cognitive development was reassessed at 7 years it was found that not only did poorer cognitive performance persist in the children with cleft-lip, but that interaction quality at 2-months accounted for the difference between groups, highlighting the long-term impact of environmental influences at this early stage (Hentges et al., 2011).

A key set of studies from the Bucharest Early Intervention Project showed that children raised in institutions suffer from a number of neuropsychological and behavioural impairments compared to never-institutionalised children. These institutions are typically characterised by low caregiver-to-child ratios, unresponsive caregiving routines, and low levels of sensory, cognitive and linguistic stimulation and so provide valuable insight into the impact of early caregiving quality. Nelson et al. (2007) investigated whether there was a sensitive period in a child's life which, if missed, made recovery from cognitive delay significantly more difficult. They found that children who had been abandoned at or shortly after birth and reared in institutions showed greatly impaired cognitive performance compared to children raised in their biological families. Children who were placed in foster-care did show significant gains in cognitive function, but those who were placed before 24 months scored significantly higher on cognitive tests than those placed after 24 months. A follow-up study when the children were 8 years old found a similar pattern of results, with children placed in foster care showing significantly higher verbal comprehension scores and marginally higher IQ scores, although overall differences between the groups were not as strong as found previously (Fox, Almas, Degnan, Nelson, & Zeanah, 2011). Furthermore, another study drawn from the same sample showed that infants placed in foster care before 24 months showed improved language outcomes at 24 and 30 months, while those placed after 24 months exhibited the same severe expressive and receptive language delays as those who remained in the institution (Windsor et al., 2011). These findings suggest that there may be a sensitive period spanning the first 2 years of life within which the onset of higher quality caregiving exerts a maximal effect on cognitive development.

1.1.4: Impact on society

While early impairments obviously affect the individual, there is also a wider economic and societal impact. Zhang et al. (2019) report that while healthy cognitive development promotes positive outcomes, such as educational attainment, employment, and earnings; lower levels of cognition are more likely to be associated with higher unemployment and criminal tendencies. This viewpoint is especially pertinent in the workplace of the 21st century, according to Knudsen, Heckman, Cameron, and Shonkoff (2006), where higher-level cognitive skills such as intellectual flexibility, adaptability and complex problem solving will be vital for success. In this context *"the personal and societal burden of diminished capacity will be formidable"* (p.10161). Early intervention, Knudsen and colleagues argue, is essential to alleviating the impact of diminished cognitive function on a nation's economy.

A recent series from The Lancet, *"Early Childhood Development",* emphasises the economic costs of poor infant development and the important role early years interventions play in promoting sustainable development in low- and middle-income countries (LMICs). They estimate that the wider cost of impaired development is particularly high in LMICs, where at least 250 million children (43%) are at risk for not achieving their developmental potential (Black et al., 2017). Richter et al. (2017) calculated that for these 250 million children the result of not reaching their developmental potential will be a 26% loss in adult earnings, *"exerting a strong downward pull and trapping families in poverty"* (p.110). In light of these findings UNICEF appear to be correct in suggesting that investing in early childhood development is one of the most cost-effective approaches to building human capital and promoting sustainable development (UNICEF Statistical Snapshot, 2014).

1.1.5: Summary

Overall, findings support the existence of a sensitive period of cognitive development in the first 2-3 years of life and suggest that a significant proportion of the foundations for future development are established during this period. One of the main objectives of the UN Secretary-General's "Global Strategy for Women's, Children's and Adolescent's Health 2016-2030" is for every child to thrive and not just survive. Dua et al. (2017) state that any research agenda seeking to support this objective must have early child development at the centre. Furthermore, as illustrated above, it is not just the individual who stands to benefit from an increased understanding of how cognitive development is shaped during infancy, but also the wider society in which they are found, and this particularly true in LMICs. This project, then, which examines the role of postnatal depression in infant cognition in India, has the potential to add to the growing literature that promotes the importance of early child development in LMIC settings for both the individual and society as a whole.

1.2 Theoretical framework

Now that it has been established that infancy is a potentially sensitive developmental window, it is important to consider how development occurs during this time. The following section will outline a theoretical framework for child development followed by how specific risk factors, such as postnatal depression, may influence development in the context of that framework.

1.2.1 Unified theory of development

The current project is guided in part by Sameroff's (2010) Unified Theory of Development, a model of developmental change within individuals throughout the life course. As the name

implies, the framework draws together a number of models and integrates them to form a unified theory of development. Whether specific risk factors will have any effect on development is, of course, dependent on how development actually occurs. Thus, the Unified Theory provides the basis for the expectation that there will be an association between maternal depression and cognitive outcomes.

1.2.1.1: Regulational model

Development and adaptation are promoted within the self through the process of *regulation***,** both within the individual *(self-regulation)* and externally to the individual *(other-regulation). Self*-regulation only occurs within the context of a dynamic relationship with experience and context, thus emphasising the role of *other*-regulation. Other-regulation has an active role in shaping adaptation and can be understood within the *contextual* model.

1.2.1.2: Contextual model

The contextual model describes the multiple sources of experience that augment or constrain individual development via other-regulation. Families, schools and neighbourhoods, and the individuals within those contexts, combine to affect the child's developmental progress. Initially, regulation is heavily supported, and sometimes exclusively provided, by caregivers. However, as the child develops, the balance between *other*-regulation and *self*-regulation shifts and the child is able to take on more responsibility for his or her own wellbeing. Thus, the capacity for self-regulation arises through the actions of others found in the various settings of the social ecology. Taken together, the psychological and biological domains of the self-system and the overarching *other*-system in which they are placed comprise the biopsychosocial aspects of the individual in context.

1.2.1.3: Personal change model

The personal change model explains the progression of competencies from infancy onwards. Two models are considered, growth and developmental. A *growth* perspective considers that all the necessary parts for development are already present in the individual and that it is their interactions that produce quantitative change in biopsychosocial aspects, without changing their interrelationships. External experience is still relevant, but only as nutrition for the inherent maturation process. The *developmental perspective* views change as a *stage* process in which the individual goes through qualitative shifts in organisation resulting in changing relations among the biopsychosocial aspects. For example, in moving

from novice to expert, the individual does not simply get better at doing things, there is a change in how they do it. Unlike the growth model, in which we would expect to observe quite gradual but consistent change, in a developmental model we would expect to see extended periods of equilibrium and relatively little change interspersed with short periods of rapid change brought on by some changes, or "punctuations", either internally, within the individual, or externally, within the social ecology. These punctuations continue to drive developmental change in the individual throughout infancy, childhood, adolescence, and into adulthood.

1.2.2: Implications for the current project

The unified model illustrates how *"biological, psychological, and social experiences foster and transform each other to explain both adaptive and maladaptive functioning across the life course"* (Sameroff, 2010, p.20). The key implication of the unified model is that researchers should start any investigation with the awareness that they are examining only a part of a large whole consisting of multiple dynamic systems. Therefore, while this project is principally interested in the relationship between maternal postnatal depression and infant cognitive development, it is understood that, while potentially significant, the role of maternal depression in cognitive development is only a small part of a much larger picture. While it is beyond the scope of this study to provide a completely comprehensive examination, the current project will follow Sameroff's (2010) recommendation to take a top-down approach by identifying and including a number of key factors that may play a role in this complex and dynamic relationship.

Another key implication of this framework is that *other-regulation* is extremely important in the early stages of development, and that often, this regulation is provided almost exclusively by the mother. Accordingly, this project focus on how maternal depression and other associated factors may relate to a mother's ability to provide the *other-regulation* required to promote the development of the child's self-regulation and, consequently, their cognitive development.

1.3 The role of maternal depression

1.3.1 Purpose of the review

This chapter will review what is known regarding the association between postnatal depression and infant cognitive development in high-income countries (HICs). The postnatal period has been emphasised in research as it represents a dual-risk in terms of their being a higher risk for mothers of experiencing depression and a higher risk for the infants in terms

of vulnerability to adverse environments (Evans et al., 2012). However, as the importance of the first 2-3 years of a child's life is being increasingly emphasised, it is also pertinent to consider the impact of extended exposure to maternal depression throughout this period. Thus, for the purposes of this study, exposure to maternal depression during infancy will be broadly delineated into two categories, early postnatal exposure and chronic exposure.

By comparing the consistency of findings regarding early postnatal and chronic effects, it will be possible to identify whether efforts to intervene on behalf of the child should be primarily focused during the first year of life, or whether efforts should be more evenly spread across infancy. First, early postnatal and chronic exposure to maternal depression will be defined. Following this, findings regarding the effects of both on cognitive development will be reviewed. This will provide a backdrop against which to discuss what is known in low- and middle-income countries (LMICs) and how the literature in these settings can progress.

1.3.2 Definitions of postnatal and chronic depression.

While the association between postnatal depression and cognitive development has been studied extensively, there is a lack of clarity and consistency in how the postnatal period is defined. Definitions vary widely in the literature, from one week to one year after delivery (Surkan, Patel, & Rahman, 2016). This thesis defines the postnatal period as anytime from birth until the infant is one year old. Chronic exposure is also inconsistently defined in the literature, as will be discussed later. Therefore, in order to capture the relevant studies, and for the sake of making straight-forward comparisons, chronic exposure is defined as exposure to depressive symptoms which begin in the postnatal period and endure throughout infancy.

Historically, there have also been large inconsistencies in what exactly constitutes postnatal depression. However, in the mid-90's O'Hara and Swain (1996) reported that a broad consensus had emerged in research to use a set of standardised diagnostic criteria. This includes symptoms such as sleep, appetite, or psychomotor disturbance, fatigue, excessive guilt, and suicidal thoughts. These symptoms should, in turn, result in some impairment in the mother's functioning. Additionally, maternal depression can be defined using self-report symptom or screening questionnaires. In these cases, depression is defined in terms of number and severity of symptoms, with women generally being diagnosed as depressed if they reach a standard cut-off score. While not necessarily meeting the criteria for a

clinical diagnosis, women endorsing high levels of symptoms often experience significant impairment in day-to-day functioning (Johnson, Weissman, & Klerman, 1992).

Postnatal depression (PND) is thus defined as a non-psychotic depressive episode of mild to major severity that occurs up to 12 months post-birth, as diagnosed by a standardised diagnostic interview, or by scoring above threshold on a validated symptom questionnaire or screening tool (Austin, 2014; Gelaye, Rondon, Araya, & Williams, 2016). Studies which have examined scores on such measures in dimensional terms, without imposing a dichotomy, will also be included (Bjelland et al., 2009). Chronic depression is defined in the same way but exposure must extend beyond the first year.

1.4 Review of evidence regarding the relationship between maternal depression and infant cognitive development

1.4.1: Why maternal depression?

The postnatal period is one of increased risk for poor maternal mental health, particularly in low- and middle-income countries (LMICs). The global prevalence of perinatal depression has been estimated at 11.9%, but a recent meta-regression showed a significant difference between high income countries (HICS) and LMICs (Woody, Ferrari, Siskind, Whiteford, & Harris, 2017). According to their study, the prevalence of postnatal depression is 9.5% in HICs and 18.7% in LMICs. These findings are supported by Fisher et al. (2012) who also report that the prevalence of common perinatal mental disorders in LMICs is higher than that found in HICs, with a weighted mean prevalence reported as 15.9% antenatally and 19.8% postnatally. Both studies drew data from studies that used a combination of diagnostic instruments and symptoms scales.

Alongside this increased risk to the mother, significant deleterious impacts of maternal depression experienced during infancy have been well documented in various domains of infant development (Hoffman, Dunn, & Njoroge, 2017; Wachs, Black, & Engle, 2009). More specifically, there is a significant body of evidence supporting an association between maternal depression and poor infant cognitive outcomes, including a number of systematic reviews and meta-analyses (Beck, 1998; Grace, Evindar, & Stewart, 2003; Kingston, McDonald, Austin, & Tough, 2015; Kingston, Tough, & Whitfield, 2012; Liu et al., 2017). Evidence from low- and middle-income countries (LMICs) is more limited. Hence, the focus of the first part of this thesis is to examine the role of maternal depression in predicting infant cognitive outcomes in LMICs, comparing it to existing evidence from HICs, and

addressing specific measurement issues that inevitably arise when making these cross-cultural comparisons.

1.4.2: Identifying Independent Effects

A key consideration when assessing the quality of the literature is whether studies have adequately controlled for the effects of depression at other time points. Two criteria need to be applied to establish whether there is an independent effect of depression within the postnatal period. Firstly, research has demonstrated that depression during the antenatal and postnatal periods exert independent effects on development and act on development through different mechanisms, meaning that it is important to distinguish between effects in each period (Osborne et al., 2018). Research also shows that the presence of a depressive episode in the postnatal period is significantly associated with a depressive episode in the antenatal period. Taken together, this emphasizes the need for studies to adequately control for the effects of antenatal depression (Norhayati, Nik Hazlina, Asrenee, & Wan Emilin, 2015; Rahman & Creed, 2007). Secondly, studies looking at the influence of a single episode of depression should account for the effects of subsequent and concurrent depressive symptoms. This is most important when both maternal depression and cognition are assessed via maternal report due to the risk of shared method variance. However, as postnatal depression significantly increases the likelihood of future episodes of depression it is also possible that any apparent cognitive impairments may be the result of subsequent episodes (Goodman, 2007; Howard et al., 2014; Richters, 1989). These two criteria will be considered throughout this review of research conducted in HICs.

1.4.3 Evidence of main effects

1.4.3.1 Evidence from systematic reviews/meta-analyses

There are a number of meta-analyses and systematic reviews that have found a significant, albeit small, effect of postnatal depression on cognitive development. Beck (1998) completed a meta-analysis of studies published from 1974 to 1995 to investigate the adverse effect on children aged 1-14 years of living with a mother with postnatal depression, revealing a small but significant overall effect across 9 studies ($r = .22$), and a smaller but still significant effect in the 4 studies that controlled for concurrent depression ($r = .17$). None of the studies were reported as having controlled for prenatal depression. There was only a marginal change when the analysis was restricted to studies of children up to 18 months old ($r = .24$). It should be noted, though, that the strength of the effect appears to grow weaker as children get older.

A similar pattern was observed by Liu et al. (2017) who performed a meta-analysis of 14 studies published between 1986 and 2013. There was no initial restriction placed on the timing of depressive symptoms with assessments ranging from the prenatal period to 42 months post-birth. Outcomes were measured in children aged 0 to 7 years. When the meta-analysis was restricted to exclude prenatal studies there was a significant unadjusted effect of maternal depressive symptoms (d = 0.27), which increased in magnitude when restricted further to studies assessing depression at 6-8 weeks post-birth (d = 0.40). Meta-analysis of the three studies that reported multivariate-adjusted linear coefficients and assessed depression in the first 6-8 weeks found a significant effect, indicating a difference of 4.17 points in the mean score of the mental development index (MDI) of the Bayley Scales of Infant Development (BSID) (Bayley, 1993, 2006). There is no mention of controlling for either prenatal or concurrent depression. Although the applicability of these findings is somewhat limited by the extended age range of developmental assessments, they do provide evidence of an overall effect of maternal depressive symptoms and indicate that this effect may be stronger from exposure in the early postnatal stages.

Another systematic review by Kingston et al. (2015) examined the association between maternal postnatal distress and toddler cognitive development. Six studies were retrieved after searching the literature from 1990-2014 that considered the impact of postnatal distress on cognitive development from 13-36 months. Five out of the six studies assessed postnatal depression, and of these, four reported a significant association with cognitive development. Each postnatal study was of medium quality, although the majority did not control for prenatal depression. The authors noted that the two studies looking at the chronicity of depressive symptoms reported contrasting findings and suggest this may be the result of different levels of risk within the samples being studied.

1.4.4 Is there a "timing effect" of early postnatal depression?

For the purposes of the current review, a timing effect is considered to be clearly present where there is a significant impact on cognitive development of a single episode of maternal depression occurring in the first 12 months post-birth and where analyses have been adjusted for prenatal depression. Although not as essential when considering observed cognitive outcomes, credit will also be given to studies which control for subsequent or concurrent depression.

1.4.4.1 Section structure

This section will be split by age of outcome in the first instance, with the primary focus being on studies which have looked at cognitive outcomes at 1 year, considered to be a short-term effect, and studies that have looked at cognitive outcomes from 1-3 years, considered to be a long-term effect. Later cognitive outcomes will be considered where relevant to the study of the impact of postnatal depression.

1.4.4.2 One year outcomes (prospective)

In reality, it is very rare to find a study that has adjusted for the effects of both prenatal and concurrent depression but there are a number of studies which have attempted to control at least one. Perra, Phillips, Fyfield, Waters, and Hay (2015) assessed the impact of postnatal depression in the first 6 months on two imitation tasks at 12 months, finding a significant effect of PND on performance in both the instrumental and arbitrary action tasks, after controlling for several confounders including mothers' past history of depressive illness, socioeconomic adversity, and infant gender. Alongside the removal of prenatal effects from the analysis, further strengths of this study include the use of a diagnostic interview for depression and a relatively large sample (n = 332).

Smith-Nielsen, Tharner, Krogh, and Vaever (2016) also assessed PND using a clinical interview in a relatively well-resourced and highly educated sample, this time at 4 months. While they did not control for prenatal depression, they did control for concurrent depression, as well as co-morbid personality disorder and infant gender. Cognitive development was assessed using the Bayley Scales of Infant Development – 3rd Edition (BSID-III) at 4 and 13 months in infants of 28 clinically depressed mothers and 55 non-depressed mothers. The authors reported a significant, negative concurrent effect of PND on infant cognition at 4 months, but no significant differences were found at 13 months. These results suggest that PND can have a concurrent effect on an infant's cognitive performance but that this effect does not necessarily persist.

Keim et al. (2011) found that PND at 4 months had no observable negative effects on cognitive development assessed by the Mullen Scales of Early Learning (MSEL) at 12 months. Analyses were not adjusted for prenatal or concurrent depression, but other confounders included prenatal stress, maternal education, family income, trait anxiety, and infant gender. For both Smith-Nielsen et al. (2016) and Keim et al. (2011), the absence of additional stressors such as low socioeconomic status, and the presence of potential protective factors, such as the presence of the partner, may have buffered the infant

against the negative long-term effects of PND, meaning that the results may not be generalisable to more high-risk samples.

One such sample, consisting of participants from a very low socioeconomic context, was studied by Lyons-Ruth, Zoll, Connell, and Grunenbaum (1986). In this study, mothers who had been clinically referred on the basis of poor mother-infant relationship and economic and social stresses within the family (but not depression) were matched with a community sample of mothers from the same neighbourhoods. PND was assessed using the Centre for Epidemiological Studies Depression Scale (CES-D), a self-report screening questionnaire, between 0-12 months, while cognitive development was assessed at 12 months using the BSID-I. Results showed that PND was associated with infant cognitive scores and accounted for 11.3% of variance in performance, after controlling for maternal IQ and behaviours, including sensitivity, warmth, verbal communication and disengagement. When compared with the findings of the two previous studies, these findings help to illustrate the importance of the context in which depression is experienced and support the principles set out in the Unified Theory of Development pertaining to the importance of maintaining a focus that captures multiple elements of the child's environment. However, it should be noted that this study did not control for prenatal or concurrent depression, and the length of the follow up between the PND and cognitive development assessments is not clear. As PND data was collected up to 12 months and the BSID-I was administered at 12 months, it is possible that some of the data is cross-sectional rather than prospective.

1.4.4.3 One year outcomes (cross-sectional)

Whiffin and Gotlib (1989) found a significant and negative effect of clinically diagnosed depression at 2 months on concurrent cognitive performance using the BSID-I (n = 50), controlling for infant age and maternal education. Similarly, Bornstein, Mash, Arterberry, and Manian (2012) observed that infants of non-depressed mothers performed significantly better when discriminating between novel and familiar views of an object in an object concept task than infants of depressed mothers at 5 months (n = 36). No covariates were included in the final analyses but participants were matched on SES, maternal age, and parity.

In more mixed findings, Kaplan et al. (2014) report no differences in performance on the cognitive and receptive language scales of BSID-III between depressed and non-depressed groups at 12 months but did find that maternal depression predicted poorer performance on the expressive language scale, after controlling for child gender and parity. Stanley,

Murray, and Stein (2004) also found no effect of maternal depression, assessed at 3 and 6 months, on instrumental learning at 3 months, after controlling for infant gender and social class. Finally, Hanley, Brain, and Oberlander (2013) found no concurrent effects of depression on either the cognitive or language subscales of the BSID-III at 10 months, controlling for maternal alcohol use and smoking during pregnancy.

Thus, the evidence for a concurrent association between early postnatal depression and cognitive development at 12 months is mixed. A notable similarity with each of these studies is the relatively small sample size. It is possible that the potentially underpowered nature of these studies has added to the mixed findings. Additionally, where significant effects have been found, their generalisability is questionable. Furthermore, none of the studies described control for the effects of prenatal depression.

1.4.4.4 Three year outcomes (prospective)

Studies that explore the effects of a single episode of depression in the first 12 months on cognitive outcomes up to 3 years are arguably more valuable as they provide insight regarding the prospective long-term effects of PND. Murray (1992) explored developmental outcomes in a low-risk sample in the UK. All mothers were screened for depression using the EPDS and then had their diagnosis confirmed by structured clinical interview at 2-3 months. Depressive history was also assessed using the Schedule for Affective Disorders and Schizophrenia. Outcomes were assessed at 9, 18 and 36 months using a range of measures. Murray found that infants of mothers with PND and no past history of depression performed significantly worse on Piaget's Object Concept Task IV at 9 months compared to infants of mothers with no depressive symptoms at 2-3 months, mothers with no past history of depression, and mothers with PND and a history of depression. Thus, it seems that depression has a greater impact on the child when the mother's first experience of the illness is during the postnatal period. One possible explanation is that in these cases, the depression is focused on the infant and the role of the mother. However, for the Stage V task at 18 months, both the PND and the PND plus depressive history groups exhibited impaired performance. In most cases, depression had remitted by 3-4 months postpartum, suggesting that depression in the early postnatal period, although brief, can have an enduring effect. Conversely, there was no effect of depression on the Stage VI task, BSID or Reynell Scales of Language Development at 18 months, although the presence of PND did appear to exacerbate the effects of social class and gender on cognitive outcomes. Concurrent depression was assessed but was not related to outcomes and so was not included in final analyses.

The sample recruited in Murray (1992) has been the subject of several studies reviewed as part of this project and will be referred to as the Cambridge sample. Although, representative of the area of recruitment, the area itself is one of considerable affluence, with 60% of the sample falling into the top 3 social classes. It is possible that the low-risk sample in this study may have been protected against some of the adverse effects of PND, leading to the null finding regarding a main effect on BSID scores at 18 months. This is exemplified in the finding that PND and social adversity interacted to predict poorer outcomes, illustrating the importance of considering the context in which depression occurs.

In a more mixed population, representing a moderate level of risk, Conroy et al. (2012) found that clinically diagnosed depression at 2 months predicted scores on the MDI of the BSID-II at 18 months, after controlling for concurrent depression and other confounders such as infant gender, maternal sensitivity and occupational status. This study, utilising a robust methodology and large, representative community sample, provides relatively strong evidence for the presence of a timing effect, although some of the effects may have been due to non-assessed prenatal depression. Similarly, a large, population-based pregnancy cohort from Greece (n = 470) revealed a significant association between PND detected using the EPDS at 8 weeks and scores on the cognitive subscale of the BSID-III at 18 months after removing mothers who reported high levels of depression during pregnancy from the analysis (Koutra et al., 2013). More recently, Kawai et al. (2017) investigated the effects of self-reported postnatal depression at 4 and 10 weeks on mother-reported non-verbal communication at 14 months. After controlling for maternal history of depression and any subsequent depressive symptoms up to 13 months, the authors report that a high EPDS score at 4 weeks was significantly and inversely associated with early gesture scores. Depression at 10 weeks did not predict non-verbal communication, suggesting that depression in the very early postpartum period may be more detrimental to development. Similarly to Murray (1992), Piteo, Yelland, and Makrides (2012) did not find any association between postnatal depression, assessed by self-report at 6 weeks and 6 months, and cognitive or language scores on the BSID-III at 18 months in a predominantly Caucasian sample from South Australia. Only a small number of mothers reported depression at both timepoints, meaning that the study lacked the statistical power to investigate the effects of extended exposure to depression. The authors did not control for concurrent depression but did include a history of diagnoses for depression, along with

maternal education, maternal social support, infant feeding, and a measure of stimulation in the home environment.

Kiernen and Huerta (2008) found that depression at 9 months was not related to infant scores on the Bracken Basic Concept Scale at 3 years, even though this was a high-risk sample with high rates of economic deprivation. The authors constructed a latent variable that drew from three measures; maternal report of feeling sad or low for 2 weeks following the child's birth, doctor diagnosed depression, and a shortened 9-item version of the Rutter Malaise Inventory. It is possible that this approach may have classified mothers with relatively low levels of depressive symptoms as depressed, which may have contributed to the non-significant findings. The authors did not control for either prenatal or concurrent depression.

Finally, Stein et al. (2008) investigated the association between maternal depression at 3, 10 and 36 months, measured using the EPDS, and verbal comprehension and expression at 36 months. Unadjusted analyses revealed significant associations between depression and language scores, at 3 and 10, but not 36, months. However, adjusted models showed that language was not predicted by depression at any time-point. Adjustment variables did not include prenatal depression.

1.4.4.5 Timing summary

As is clear from the studies discussed above, evidence is mixed regarding the influence early postnatal depression on cognitive development. Out of the 11 prospective studies, eight controlled for either prenatal or concurrent depression but only two studies controlled for both. Further, of those that controlled for prenatal depression, the majority only controlled for general depressive history before childbirth. This is not ideal as this does not distinguish prenatal depression from more historical episodes and the methods used to record this data are often retrospective. As prenatal depression is understood to exert specific biological and physiological effects on the developing foetus, it is important to be able to isolate and control for this specific mechanism of effect. Future studies should endeavour to directly assess prenatal depression during pregnancy.

With that caveat, there is some evidence of an effect of early postnatal depression on cognitive development. Four out of six studies looking at outcomes beyond the first year found that postnatal depression predicted poorer cognitive outcomes during the child's second year, two of which controlled for historical and current depression, while two out of four studies found a prospective effect in the first year of the child's life. While this is

certainly enough to suggest that there is some relationship between early postnatal depression and cognitive development, the number of studies which did not find a significant relationship is not insignificant either. This inconsistency means that it is not appropriate to state conclusively that there is a timing effect of postnatal depression. More discussion is required to explore other factors that may influence any association at different time-points.

1.4.5 Chronic exposure to maternal depressive symptoms throughout infancy

Before proceeding with the review, it is important to draw attention to a common issue in the study of chronic exposure. While many studies report the apparent effects of chronicity, the definition, conceptualisation, and measurement of the chronic factor varies widely, and not all are as chronic as they first appear. This is illustrated in the following study. Petterson and Albers (2001) assessed maternal depression in a large, culturally diverse national sample (n = 7677) during pregnancy and at 28-50 months. Mothers were considered to be experiencing chronic depression if they reported symptoms above the cut-off at both time-points. While there was a significant effect of depression in the prenatal period only, depression at both timepoints resulted in a much more substantial deficit in cognitive development, for both boys and girls. This approach is significantly limited, however, in that there is a large gap between the first and second waves of assessment and no indication of whether mothers in the chronically depressed group remained depressed throughout that time. As the depression instrument used only asked about symptoms experienced in the past week, it may be misleading to classify a mother as chronically depressed based on reports of how she felt during two weeks of what may have been a four-year period. Indeed, it is possible that mothers who were not depressed at either time-point may have experienced an extended period of depression in-between assessments which was not identified.

These concerns are validated by a systematic review of longitudinal studies on antenatal and postnatal depression which found that a substantial number of mothers moved between the groups categorised as depressed or non-depressed throughout follow-up (Underwood, Waldie, D'Souza, Peterson, & Morton, 2016). This highlights the importance of ensuring that studies investigating the longitudinal effects of chronic depression collect data at sufficiently regular intervals and then analyse it in a way that maximises the data obtained. It should also be noted that while adherence to the requirements of adjusting for the effects of prenatal and concurrent depressive symptoms will still be reviewed, many studies include the concurrent time-point in their chronic variable.

1.4.5.1 Chronic exposure assessed by grouping multiple assessments

Azak (2012) investigated how maternal depression diagnosed at 6 months and shown to be stable across follow-up influenced the trajectory of cognitive development from 6-18 months. A small sample of mothers who had been referred for treatment for postpartum mental illness were compared to a group of healthy controls. Mothers were first screened using the CES-D and those who scored above the threshold had their diagnosis confirmed using the MINI diagnostic interview. Follow-up assessments at 12 and 18 months showed that on average mothers in the depressed group remained depressed, whereas mothers in the non-depressed group remained non-depressed, and that depression exerted a stable effect on infants' trajectories of cognitive development. In other words, cognitive development, as measured by the Mullen Scales of Early Learning, was significantly lower for infants in the depressed group at 6 months and remained lower through to 18 months. However, the majority of depressed mothers did also report experiencing prenatal depression, and this may have also contributed to poorer cognitive outcomes in the depressed group.

Sutter-Dalley et al. (2012) also assessed cognitive development at multiple time-points, from 3-24 months. The analysis was designed to tease apart the independent effect of a single episode of postpartum depression from effects of chronic and concurrent depression. Initial results indicated that the severity of depression 6 weeks after delivery, as measured by the EPDS, predicted lower child cognitive performance over the follow-up. However, after adjusting for subsequent episodes of depression, the initial association was reduced to the level of a trend. These findings demonstrate the importance of accounting for subsequent episodes of maternal depression and signal that it may be inappropriate to draw concrete conclusions regarding the effect of the timing of maternal depression without doing so.

Milgrom, Westley, and Gemmill (2004) recruited a sample of depressed mothers who had recently given birth and were being treated as inpatients at mother-baby psychiatric units, along with a control group of well mothers from community health centres. Depressed mothers had been diagnosed with a major depressive disorder remained at least moderately depressed at 15 weeks, and then at 12, 24 and 42 months. Results show that infants of depressed mothers scored significantly lower on both the Wechsler Preschool and Primary Scale of Intelligence and the overall Cognitive/Language profile of the Early Screening Profile at 42 months. Similarly to Azak (2012) and Sutter-Dalley et al. (2012), this study shows that relatively stable depression starting in the postnatal period and persisting

throughout infancy may be having a significant effect on cognitive outcomes. Again though, the lack of control for prenatal symptoms mean that it is possible that depression during that period was at least partially responsible for the observed effect.

A key population study was carried out in the US by the National Institute of Child Health and Development Early Child Care Research Network (NICHD ECRN, 1999). This study explored the link between chronic maternal depression, assessed at 5 time-points from birth to 3 years, and child functioning at 3 years. The sample was split into 3 groups, never depressed, sometimes depressed, and chronically depressed. Mothers were allocated to the *chronically depressed* group if they reported elevated depressive CES-D scores at four or five assessments. Although there was a significant harmful effect for both the chronic and sometimes depressed groups, the effect was stronger for children of chronically depressed mothers, who exhibited deficits in school readiness, verbal comprehension, and expressive language. Results were not adjusted for the influence of prenatal depression and the concurrent time-point was included in the construction of different groups.

In a relatively low-risk sample, Cornish et al. (2005), found that depression reported during the first 4 months but that had resolved by 12 months postpartum had no effect on cognitive or language development. In contrast, chronic depression, lasting throughout 12 months and beyond, was associated with lowered cognitive, but not language, performance in 15-month-old infants. Depression was assessed using the Composite International Diagnostic Interview (CIDI) at 4 and 12 months, and also by the CES-D at 12 and 15 months. At 12 months, the CIDI was utilised in conjunction with additional criteria to establish whether continuing or recurrent depression had occurred between assessments.

Sharp et al. (1995) recruited women from two socio-economically deprived areas of London and assessed depression at 3 and 12 months using the Clinical Interview Schedule, and then retrospectively for the intervening time at 4 years using the Schedule for Affective Disorders. Interestingly, while depression in the first year was related to lower cognitive scores at 4 years, it made no difference if the mothers reported additional depression from 1-4 years or not. Moreover, depression during the 1-4 year period without a preceding episode in the first year, was not related to cognitive performance. It is possible that the lack of effect regarding chronicity was due to differences in how depression was assessed at each time point, with retrospective reporting likely to be less reliable. Planned contrasts revealed that the effect of depression in the first postnatal year was confined to boys.

Finally, Brennan et al. (2000) investigated the effects of chronicity, severity and timing in a large and relatively low SES Australian sample (n = 4953). Mothers' depression was assessed using the Delusions-Symptoms-States Inventory during pregnancy, a few days postpartum, and at 6 months and 5 years. Both severity and chronicity of depressive symptoms were related to poorer scores on the Peabody Picture Vocabulary Test (PPVT) at 5 years. Differences predicted by severe or moderate depression reported at any one time point were not significantly different from depression reported at any other one timepoint. These findings suggest that no specific time-point, prenatal or postnatal, is more important than another for development, but that exposure to severe or prolonged depression is disadvantageous to the infant.

A limitation of this study is the large time gap between the 6 month and 5-year depression assessments, wherein the mental wellbeing of mothers may have varied considerably. A related issue, regarding the definition of chronic depression, is that chronicity includes both prenatal and concurrent depression. While the timing analysis suggests that there is no difference in impact of depression at different points, it is possible these time-points may interact with others to produce differences that are not picked up in the current analysis. While a number of studies have modelled trajectories of maternal depression across childhood, there appears to be only one which has modelled trajectories throughout infancy in relation to child cognitive outcomes. Park, Brain, Grunau, Diamond, and Oberlander (2018) investigated the influence of maternal depression on child executive function at 6 years in a sample of mothers at high risk for mental disorder. While this age goes beyond the focus of the current study, it usefully illustrates the use of trajectories in this area.

Depression was assessed using a combination of self-report measures during the second and third trimesters of pregnancy, and then post-birth at 6 weeks, 3 months, 6 months, 10 months, 3 years, and 6 years of age. Trajectories of depressive symptoms were produced using growth mixture modelling with standardised scores from each measure up to 3 years. This produced a model with three trajectories of maternal depressive symptoms. The three groups were as follows: low, characterised by mothers with consistently low symptom scores; increasing, characterised by mothers who exhibited moderate symptoms during pregnancy and increasing symptomology over time; and high, characterised by mothers with high levels of symptoms during pregnancy that decreased over time. The key finding was that children of mothers in the low and decreasing groups showed similar levels of executive function, while the children of mothers in the increasing group had poorer

executive function. Importantly, the authors also controlled for the effects of maternal depression at 6 years, meaning that results can be interpreted to suggest that a prolonged and increasing exposure to maternal depression starting in pregnancy and continuing throughout the first 3 years of life can have a persistent effect on executive function in later childhood. Another interesting observation is that even though mothers in the decreasing group showed persistently higher symptomology than mothers in the low group throughout follow-up, there were no significant differences in executive function. This suggests that if mothers who initially exhibit high levels of depression symptoms experience even partial remission during the first 3 years of their child's lives, the child's cognitive development remains unaffected.

These findings illustrate the benefits of modelling trajectories of maternal depression in terms of exploring the heterogeneity of different mothers' experiences, though the generalisability of the findings is limited somewhat by the small sample size. Further, the inclusion of prenatal symptoms in the overall trajectories could also be viewed as a limitation if one was being strict about the criteria for establishing the presence of a postnatal effect. However, the authors note that research has shown that postnatal depression may be a continuation of prenatal depression (Underwood et al. 2016). Possibly for this reason, prenatal depression was included in the development of trajectories, rather than as an adjustment variable. While this would not be the approach used in this thesis, the authors have given relatively clear justification for this decision.

1.4.5.2 Chronic exposure summary

The findings from the studies reviewed above reveal a fairly consistent pattern whereby chronic maternal depression during the first few years of a child's life is observed to have a significant detrimental effect on cognitive development. However, it is also apparent that there are many inconsistencies in how studies are designed and carried out. This is most significantly seen in how *chronic depression* is defined and measured. While some studies assess depression at either end of a long time-period and assume that depression was continuous throughout, others assess depression across a higher frequency of shorter intervals. Even among those who assess depression at more time-points, their definitions of what constitutes a chronic exposure vary considerably. More recent literature has advocated for the use of trajectories of maternal depression as this is thought to uncover more of the subtleties of depression and the heterogeneity of experience between different mothers. Thus, while there is substantial evidence that chronic exposure to

maternal depression does impede an infant's cognitive development, there is still a need for research that takes a consistent and robust approach to measuring chronic depression.

1.5 Postnatal depression and cognitive development – mediated and moderated effects

The theory of Unified Development (Sameroff, 2010) discussed earlier in this chapter emphasises that child development is the result of a complex interaction of different proximal and distal factors. Consideration of these factors and their interactive or confounding effects may help to explain some of the inconsistent effects that have been particularly evident thus far, particularly with regards to brief episodes of depression, which may only be significantly detrimental in the context of other risks. At this stage, a number of new studies will be introduced alongside findings regarding indirect effects from the some of the studies already reviewed. Factors include caregiving quality, infant gender, socioeconomic status, maternal IQ, maternal education, maternal age, and parity. The findings presented here will inform the analytic approach taken by this project in terms of the moderated effects to be explored and the confounding effects to be controlled for.

1.5.1 Maternal caregiving

1.5.1.1 The importance of caregiving for the development of the child.

A key factor in promoting early development is the *other-regulation* provided by external sources (Sameroff, 2010). Particularly important is the regulatory support offered by the primary caregiver, a role generally fulfilled by the mother. This caregiving environment has been broadly conceptualised as nurturing care and more specifically as constructs such as maternal sensitivity and responsiveness.

The recent *Lancet* series, *Advancing Early Childhood Development,* emphasises that children who develop in the context of nurturing care are more likely to reach their developmental potential and responsive caregiving is strongly recommended as an integral focus of early intervention (Black et al., 2017). Nurturing care consists of *"a stable environment that is sensitive to children's health and nutritional needs, with protection from threats, opportunities for early learning, and interactions that are responsive, emotionally supportive, and developmentally stimulating"* (Britto et al., 2017, p.91). These nurturing interactions are crucial in the mitigation of the adversity found in high-risk samples or LMIC settings but are liable to breakdown under conditions of extreme poverty, family conflict, and other forms of individual stress, such as maternal depression (Daelmans et al., 2017). Thus, the nurturing relationship that is most pivotal in optimising development under stressful conditions is itself at heightened risk of being compromised in those contexts.

Although there are many aspects of the caregiving environment that may contribute to child development, the concept of maternal sensitivity originally proposed by Mary Ainsworth is considered to be particularly important. Ainsworth defined maternal sensitivity as "*the mother's ability to perceive child signals, interpret these signals correctly, and to respond to these signals promptly and appropriately*" (Ainsworth, Bell, & Stayton, 1974).

1.5.1.2 How does caregiving promote development?

According to Britto et al. (2017) high quality nurturing care mediates the development of key brain regions, promotes developmental adaptations, and protects the infant from the negative effects of stress and adversity. Maternal sensitivity is fundamental to nurturing care, as sensitive and contingent responding to infant cues is thought to promote attachment security and development of the strategies needed for an infant to develop self-regulation skills, as well optimal strategies for controlling and managing behaviours (Grant, McMahon, Reilly, & Austin, 2015; Mills-Koonce, Gariepy, Sutton, & Cox, 2008). In contrast, inadequate parenting exacerbates risk by increasing exposure to early life stress, interfering with a mother's ability to meet the infant's social and emotional needs, and disrupting development of early physiological regulation (Goodman, 2007).

1.5.1.3 How does maternal depression affect caregiving?

There is a growing body of research that shows that high levels of maternal depression can compromise sensitive caregiving (Beck, 1995; Campbell, Matestic, von Stauffenberg, Mohan, & Kirchner, 2007; Crockenberg & Leerkes, 2003a; Mills Koonce et al., 2008; Murray, Fearon, & Cooper, 2015, Pearson et al., 2012). A recent meta-analysis (k = 48, n = 4934) found a significant but small aggregate effect size between depression and maternal sensitivity (r = -.16), and a medium sized effect when restricted to studies which compared a depressed and non-depressed/control group (r = -.35) (Bernard, Nissim, Vaccaro, Harris, & Lindhiem, 2018). Maternal depressive symptoms appear, then, to represent a key threat to the ability of the mother to provide the healthy interactions and learning opportunities.

In their meta-analysis, Bernard et al. (2018) suggest a number of mechanisms by which depressive symptoms may interfere with sensitive parenting, including blunted neural responses to infant cues, negative cognitions about infant crying, reduced emotion regulation capacities, and sleep disturbance and fatigue. Depressed mothers may also be cognitively predisposed to respond negatively to infant distress and less well equipped to respond positively (Murray et al., 1996). Other factors may include difficulty concentrating,

lack of interest in daily activities, or the mothers' preoccupation with their own negative thoughts and cognitions, all of which may interfere with their capacity to notice and respond contingently to their interpersonal environments (Ahun & Cote, 2018; Crockenberg & Leerkes, 2003a; Stein et al., 2008, 2012). In another review, Murray et al. (2015) state that although less substantial than high-risk samples, findings from low-risk samples show subtle differences in caregiving quality, including reductions in behavioural responsivity and sensitivity to infant cues, less physical touch and reduced overt signs of affection, and slower and less responsive speech.

Goodman & Gotlib, (1999) hypothesised that the negative maternal cognitions, behaviours, and affect that characterise depression may result in a depressed mother being an inadequate social partner who is unable to meet the social and emotional needs of the child. Field (2010) states that this depression induced inadequacy manifests in two main interaction styles and that depressed mothers were likely to display an interaction style that was either withdrawn, passive and under-stimulating or intrusive, controlling and overstimulating.

Mothers whose interactions are characterised by withdrawn, non-responsive behaviour are often not attuned to the child's emotions and behaviours and cannot, therefore, offer the 'emotional scaffolding' required to supportively contain the child's difficult emotions. In turn, this can lead to dysregulated affect in the infant which can impair attention and disrupt information retrieval. Self-regulation is also promoted via maternal positive affect in mother-child interactions, and so when this element is not present infants miss out on a key motivational system that works to facilitate cognitive development (Hanley et al., 2013). Dysregulated emotion, in turn, can lead to missed opportunities for learning and development for the infant.

Intrusive interactions are characterised by interventions that are not responsive to the infant's affect, state or interest, and are often in direct opposition to the infant's focus of attention (Swingler, Perry, Calkins, & Bell, 2017). These interactions create rapidly shifting sources of arousal and stimulation that actively interfere with the child's ability to practice effective control of attention and goal directed behaviour, both of which are important building blocks for later executive function (Gueron-Sela et al., 2018). Thus, both intrusive and withdrawn parenting styles reduce the number of opportunities for exploration and social engagement at home, putting infants at risk for not acquiring important developmental skills (Black et al., 2007).

1.5.1.4 Evidence for an interactive effect of maternal depression and caregiving on cognitive outcomes

One of the key studies in this area is the NICHD ECRN (1999) study discussed earlier. As well as the exploring the effects of chronic depression, the authors also investigated how maternal caregiving was affected by depression and its role in mediating the association between depression and cognitive outcomes. Caregiving was assessed at 6, 15, 24 and 36 months and operationalised as a composite of three subscales used to rate maternal behaviour during a 15-minute play interaction at each timepoint. Subscales were sensitivity to non-distress, positive regard and intrusiveness. This composite, labelled maternal sensitivity, was then averaged across the four time-points and included as a single variable in analyses.

As expected, there were significant differences in maternal sensitivity between depression groups. The never depressed group was significantly more sensitive than both the sometimes and chronically depressed groups. Further, the chronically depressed group exhibited a unique pattern change, being the only group to show systematic decline in sensitivity in the second year and then some recovery in the 3rd year. The authors contend that this decline may be the result of developmental change in the child at 24 months, whereby verbal ability and bids for autonomy may make play interactions more challenging. It is possible that chronically depressed mothers are particularly vulnerable to these more challenging behaviours. In addition to this, results showed that the income-to-needs ratio moderated the relationship between depression and maternal sensitivity, with depressed mothers who were less financially stable showing less sensitivity than depressed mothers who were more financially secure. It seems plausible that mothers with fewer financial worries were less preoccupied with stresses and worries than poorer women, and thus were more available to notice and respond to infant cues.

In terms of child development, maternal sensitivity was a significant predictor of all outcomes, over and above the effects of depression. Maternal sensitivity also accounted for some of the effects of depressive symptoms, with depression-group differences for school readiness, verbal comprehension and expressive language being attenuated or eliminated once the effect of maternal sensitivity was controlled in the analyses. Mediation effects, while significant, were small however, indicating that both sensitivity and depression are important in understanding the results. Finally, in the moderation analysis, there was a significant effect of the interaction between depressed group and sensitivity only for expressive language. Children of mothers in the sometimes-depressed group

performed better when their mothers were more sensitive. Thus, it appears that high levels of maternal sensitivity might act as a buffer in certain areas of development for mothers who are report being sometimes depressed. Interestingly, this effect was not found in the chronic group, suggesting that chronic depression may pose such a risk to development that even more sensitive parenting cannot protect the child from its impact.

Milgrom et al. (2004) also investigated the mediating role of maternal caregiving, conceptualised as maternal responsiveness at 6 months. Depressed mothers were significantly less responsive than non-depressed mothers. When maternal responsiveness was included as a mediator of the relationship between chronic depression and child IQ the effect of depression became non-significant, although there was no mediation evident for the relationship between depression and cognitive/language profile scores. More recently, Gueron-Sela et al. (2018), found evidence that harsh and intrusive parenting may play a similar mediating role. Depressive symptoms were assessed at 6, 15 and 24 months, and executive function (EF) was assessed at 36 and 48 months. There were significant direct effects of maternal depressive symptoms at 15 and 24 months on EF scores at 48 months. However, the effect of depression at 15 months was partially mediated by a factor representing harsh and intrusive parenting at 24 months. This is consistent with the theory that the risk posed by maternal depressive symptoms during infancy and toddlerhood is at least partially exerted through compromised parenting styles.

1.5.2 Infant gender

1.5.2.1 Sex differences in sensitivity/vulnerability to environment

While the studies discussed above found an overall effect of caregiving, there is growing evidence that there may be sex differences in how sensitive infants are to their environment (Donald et al., 2019). Although boys typically lag behind girls in early cognitive development (Gur et al., 2012; Weber, Darmstadt, & Rao, 2017) it is now thought that specific attributes of male infants and their interactions with caregivers may be putting them at a further disadvantage in the context of mental health issues such as maternal depression. At the root of this difference is thought to be an increased level of negative affectivity coupled with a relatively low self-regulatory capacity in male infants as compared to female infants (Azak, 2012). As males are less able to self-regulate their emotional and attentional states, they rely much more heavily on the facilitation and scaffolding of an emotionally healthy caregiver (Murray & Cooper, 1997), linking back to the role of other-regulation in promoting early development and adaptation emphasised in the Unified Model (Sameroff, 2010). This general maturational advantage held by female infants may

also confer an advantage in the presence of maternal depression (Grace et al., 2003). As discussed earlier, postnatal depression may impair the mother's ability to act as an adequate social partner to the child. This may present a more substantial developmental obstacle to male infants who, according to the theory above, are more needful of the emotional and attentional scaffolding less available in the context of depression. Additionally, the presence of higher levels of negative affect in male infants is likely to impose on the mother-infant relationship by providing consistently negative feedback to the mother's caregiving attempts, prompting more affectively negative responses from her in return, and lowering sensitivity over time (Bridgett et al., 2009, Cecil, Pickles, Hill, & Sharp, 2017; Crockenberg & Leerkes, 2003b). In combination, these two factors may result in a cycle of regulatory problems specific to the relationship between mothers and sons (Tronick & Reck, 2009).

1.5.2.2 Evidence for sex differences in the effect of postnatal depression on cognitive development

Strong evidence regarding the interplay between maternal depression, caregiving quality and gender was provided by the Cambridge studies discussed earlier. In further analysis of a subsample of the same cohort, Murray et al. (1996) found a main effect of gender and a significant depression by gender interaction on cognitive development. While there was no effect of postnatal depression on BSID-I scores at 18 months in the whole sample, boys of postnatally depressed women performed significantly worse on the BSID than did boys of well women and there was no difference between groups for girls. Depressed mothers were also rated as less sensitive during interactions with male infants and expressed fewer affirmations, and more negations, of their infants' behaviour, compared to their interactions with female infants. Furthermore, maternal behaviour partially accounted for the effects of the gender by depression interaction on cognitive development. This suggests that boys' cognitive development may be uniquely affected by the caregiving style of depressed mothers. This was also shown by Murray, Kempton, Woolgar, and Cooper (1993) in the same sample using a different measure of caregiving. In this study, the authors examined the role of the quality of maternal communication at 2 months in relation to PND and cognitive outcomes. The speech of depressed mothers towards male infants was less infant-focused and, when it was infant-focused, more critical and hostile toward the infant. Further analysis revealed the poor-quality maternal communication experienced by male infants largely mediated the association between depression and infant cognitive outcomes in the first 18 months. Thus, it appears that male infants may be

particularly vulnerable to the effects of postnatal depression on cognitive development via compromised maternal communication. This is consistent with the theory that male infants may elicit a more negative style of parenting from their caregivers which could in turn affect development (Crockenberg & Leerkes, 2003b).

The authors next performed a further follow-up with a subsample of participants when the children were 16 years old to examine the cumulative effects on academic performance of maternal depression (postnatal and subsequent), maternal IQ, child sex, earlier cognitive development, and mother–child interactions (Murray et al., 2010). Consistent with their previous findings, they found that boys of PND mothers were more adversely affected than girls of PND mothers and children in the control group. Neither chronic nor recent exposure to maternal depression had any significant effects on cognitive development. Rather, this male-specific effect was principally accounted for by impairments in early cognitive development. Interestingly, the authors report that evidence for an effect of maternal depression on girls' cognitive development across the same period was completely lacking. Similarly, while IQ scores of boys during childhood predicted GSCE outcomes, there was only a weak association between childhood IQ and GSCE outcome for girls. Postnatal depression was also found to have continuing negative effects on maternal interactions throughout childhood for boys but not girls. These findings suggest that the gender differences in response to postnatal depression and impaired caregiving can be very long-lasting. It is possible that the adverse effect of depression on cognitive development during childhood in boys, due at least in part to impaired interactions, served to constrain subsequent cognitive development during adolescence. Although this was a small follow-up sample, the authors conducted a set of Monte Carlo simulations which indicated that the findings were robust. Still, some caution in needed when generalising these findings.

This set of studies from Murray and colleagues provide strong evidence that PND, acting through impaired mother-infant interactions, confers a higher level of risk on male infants' cognitive development, which in turn can have long-lasting effects on academic success.

Evidence for sex differences in cognitive development in response to maternal depression outside of the context of assessed caregiving quality is also quite consistent. Infant gender was implicated in the study by Sharp et al. (1995), discussed earlier. Following the finding that depression during the first postnatal year predicted poorer cognitive outcomes, planned contrasts revealed that the impact of postnatal depression was actually bound up in an interaction with gender, and that the impact of maternal depression was reliable only

for boys. Indeed, girls appeared to be remarkably well protected against the deleterious effects of their mother's poor mental health. A follow-up of the same sample found that boys of mothers depressed in the first year postpartum, and specifically the first 3 months, remained the most at-risk group relative to girls in the same situation and children of mothers who reported depression after this point (Hay et al., 2001). Consistent with this finding, Milgrom et al. (2004) found that cognitive problems as a result of maternal depression were more pronounced in boys and Azak (2012) reported that sons of depressed mothers had the lowest growth trajectories over the follow-up.

Finally, Kurstjens & Wolke (2001) used a retrospective design to explore the relationship between different features of maternal depression within the first 7 years and child intellectual development at 20 months, 4.8 years and 6.3 years. There were no main effects of timing (reported within the 1st year postnatal), duration (longest episode up to 6 months vs longer than 6 months), severity (minor vs severe), recency (within the last year), frequency (single vs multiple episodes) or chronicity (major depression starting in 1st year and involved multiple episodes) of depression on various measures of cognitive development at each follow-up point. Although they also failed to find a significant two-way effect of depression status and infant gender, there was evidence of a more complex effect of chronic depression at 6 years. There were two 3-way interactions showing that boys who had low SES or high neonatal risk status and who were exposed to chronic depression had lower cognitive scores at 6 years. Although these interactions were not significant at either 20 months or 4.8 years, this finding speaks to the theory that long-term exposure to a combination of adversities may be detrimental for a potentially vulnerable subgroup such as male infants.

Sex differences do not enjoy the unanimous support of the literature, however. Several studies discussed in prior sections tested out the moderating role of infant gender and did not find a significant effect. Perra et al. (2015), who found a significant effect of PND at 6 months on two imitation tasks at 12 months report that this effect was not moderated by infant gender. Similarly, Smith-Nielsen et al. (2016), who found a significant cross-sectional effect of PND on BSID-III cognitive outcome at 4 months report that moderation analyses at 4 and 13 months were non-significant for both cognitive and language outcomes. Finally, Cornish et al. (2005) found that chronic depression across the first 12 months and beyond did predict significantly impaired BSID-II cognitive scores and that effects were similar for boys and girls.

Interestingly, both Cornish et al. (2005) and Smith-Nielsen et al. (2016) assessed very low risk samples. Taking these results together with Kurstjens and Wolke (2001), one might conclude that boys are at higher risk for impaired development only when maternal depression is accompanied by other risks, such as impaired caregiving or lower socioeconomic status. While the findings by Murray et al. (1992, 1993, 2010) from a relatively well-off sample in the UK speak to a level of increased risk for boys in the context of compromised caregiving, regardless of other risk factors, there is a clear trend throughout the findings reported in this review that socioeconomic risk also plays an important role.

1.6 Other covariates

1.6.1 Socioeconomic Status: how does it impact cognitive development?

As indicated above, socioeconomic status (SES) may significantly impact cognitive development. According to Letourneau, Duffet-Leger, Levac, Watson, and Young-Morris (2011), SES is widely accepted to have a negative impact on various domains of child development and is typically operationalised as a composite variable of at least two factors. Different factors may include parental education, marital status, employment status, occupations prestige, household income, or eligibility for subsidy.

One of the primary ways in which low SES can impact on cognitive development is through the reduction of resources, both tangible and intangible, available for the parents to utilise in providing a stimulating environment (Molfese, DiLalla, & Lovelace, 1996; Conger & Donnellan, 2007; Jensen, Dumontheil, & Barker, 2014). Poorer mothers often have to deal with a larger number of competing demands than more affluent mothers, who may be able to outsource some of their responsibilities to others. These extra demands within the family context reduce the resources available to the mother to respond sensitively and appropriately to her child, potentially leading to a deficit in the other-regulation required for optimal development (Crockenberg & Leerkes, 2003b, Sameroff, 2010). Low SES can also make families more vulnerable to stressors such as economic hardship or poor mental health, which can lead to a further undermining of the resources required for skilful parenting (Conger & Donnellan, 2007; Lee, 2013).

1.6.2 Evidence that SES impacts cognitive development

A number of studies have shown a main effect of SES on cognitive development and that the consequences of socioeconomic inequalities generally emerge early and, without intervention, may persist throughout the life course (Najman et al., 2004). These early

differences are demonstrated by Feinstein (2003), who reported a 13-percentile gap in an index of cognitive development at 22 months between children from high and low SES families. Similarly, findings from the UK Millennium Cohort revealed a significant socioeconomic gradient in children's development by 3 years, after controlling for a range of parental and child characteristics (Doyle et al., 2009).

A recent meta-analysis reported very small to small significant effects of SES on various indicators of child development, including early language development (Letourneau et al., 2011). The authors reported that significant relationships were almost always explained by some combination of additional factors and agree with the conclusions of another meta-analysis (Sirin, 2005) that the association between SES and development or academic achievement is generally moderated by multiple factors at an individual, familial and community level. These findings are echoed by Tong, Baghurst, Vimpani, and McMichael (2007) who found that although socioeconomic position had an independent effect on cognitive function throughout childhood, this relationship was markedly attenuated by controlling for maternal IQ and the quality of the home environment.

Stein et al. (2008) discuss the association of postnatal depression and cognitive development in the context of impaired caregiving and low SES. While maternal caregiving is considered to be a critical factor in the transmission of risk, not all studies have found the link between postnatal depression and caregiving quality. Indeed, there is a significant body of research that suggests that the theorised impact of depression on caregiving, and consequently on development, is at least partially dependent on socioeconomic factors (Lovejoy, Graczyk, O'Hare, & Neuman, 2000, NICHD ECRN, 1999). In their own study, Stein and colleagues reported that the adverse effect of postnatal depression on caregiving quality was stronger in lower SES families but that there was an independent effect of caregiving in both high and low SES groups. These findings suggest that while maternal caregiving may be particularly vulnerable to postnatal depression in lower SES settings, the subsequent relationship between caregiving and development might not be moderated in the same way.

Overall, the evidence points to a significant role for SES in the relationship between postnatal depression, maternal caregiving quality, and infant cognitive development. While not fully clear, the findings discussed above reveal another layer of the complex interplay of factors that may promote or inhibit child development and highlight the importance of considering the role of SES in future research.

1.6.3 Maternal education and IQ

Although sometimes factored into measures of SES, maternal education, IQ and age are important independent indicators of infant cognitive development and so will also be considered briefly. As indicated in the discussion of SES, parental, and particularly maternal, education appears to play an important role in child development. Better educated parents are, in general, better equipped to provide nurturing care (Black et al., 2017) and facilitate stimulating home environments (Doyle et al., 2009). Higher levels of maternal education can also promote child development through reduced risk of maternal depression, better child nutritional status, greater knowledge about child development, having higher education aspirations for children, and being better able to access and benefit from interventions (Walker et al., 2011). Maternal IQ and education are also important indicators of the known substantial genetic variation in cognition (Bouchard & McGue, 1981; Davies et al., 2011).

1.6.4 Maternal age

Older maternal age appears to show a fairly consistent positive influence on offspring cognitive outcomes. This effect has been shown throughout infancy and early childhood (Saha et al., 2009), middle childhood (Fergusson & Lynskey, 1993; Goisis, Schneider, & Myrskyla, 2017), and early adulthood (Fegusson & Woodward, 1999). Maternal age is thought to exert its effect through association with a large number of psychosocial variables such as higher socioeconomic status and increased educational attainment, which have themselves been identified as independent predictors of development (Tearne, 2015). Additionally, Barnes, Gardiner, Sutcliffe, and Melhuish (2014) found that positive and responsive parenting increased with age, while household chaos and harsh discipline decreased with age.

1.6.5 Parity

While there is a rich literature exploring the effects of birth order on development, much of this research has been criticised for using flawed cross-sectional, between-family designs which do not take account of potentially confounding differences between families, such as parental IQ or maternal education (Rodgers, 2014). More recent studies have failed to find what was previously considered a well-established effect and it has been suggested that the association between birth order and development is a methodological illusion (Rodgers, Cleveland, van den Oord, & Rowe, 2000).

There is some debate over the existence of birth-order effects, with some now arguing that the previously well-established evidence was actually the product of poor methodology (Rodgers, 2014). However, others have suggested that one reason recent findings have been non-significant is that the actual effect is small and that many studies have been statistically underpowered (Bjerkdal, Kristensen, Skjeret, & Brevik, 2007). Accordingly, a novel study by Kristensen and Bjerkedal (2007) investigated a birth-order effect that incorporated biological and social ranks within the family using a large sample drawn from Norwegian conscription data. They hypothesised a social interaction effect whereby second-born individuals who lost an older sibling, and were thus promoted to first-born rank, would score more highly than individuals with a biological or social rank of second-born. Their findings showed that this was indeed the case, that IQ scores were negatively associated with both birth order and social order, and that in adjusted models, the effect of biological order was no longer significant after accounting for social order. They argue that this provides evidence that the relation between birth order and IQ is dependent on the social rank in the family, and not birth order per se. Using a similar, within-family design, Barclay (2015) also found birth order effects using Swedish conscription data.

Although there are a number of theories as to why these birth-order effects may exist, one of the more prominent theories is the resource dilution hypothesis. According to this theory, lower-birth-order children have an advantage in terms of access to parental resource and investment, while higher birth order children may experience decreasing amounts of interaction and stimulation from caregivers (Heiland, 2009). In terms of the unified model, this effect could be explained in terms of later-born children having reduced access to the resource of other-regulation provided by parents.

1.6.6 Nutrition

Overall, infant malnutrition represents a significant threat to child development, and remains a serious challenge in many developing countries. Black et al. (2017) reported that, in 2010, 249.4 million children under the age of 5 were at risk of not achieving their developmental potential due to extreme poverty and stunting.

A number of factors have an important role in promoting healthy brain development through adequate nutrition. Maternal nutrition is particularly important during pregnancy when sufficient energy and protein supply, essential fatty acids, and various key micronutrients are required to facilitate and coordinate normal brain development. Poor nutrition, ranging from extreme cases of famine to specific nutrient deficiencies or dietary

imbalances, can directly impair brain development and result in adverse neurodevelopmental outcomes (Lindsay, Buss, Wadhwa, & Entringer, 2019). In addition to observed foetal programming effects on the brain, poor nutrition can also lead to intrauterine growth restriction, premature birth and low-birth weight (LBW), each of which has been linked to impaired cognitive development (Britto et al., 2017). LBW appears to be particularly important, with evidence from both HICs and LMICs demonstrating a significant association with infant neurodevelopment. One study in the US found that extremely pre-term or extremely LBW infants showed cognitive deficits 3 to 6 times higher than normal birthweight children at 6 years on a comprehensive battery of neurodevelopmental tests (Orchinik et al., 2011), while another study in Sweden found that very LBW children demonstrated deficits in reading skills and cognitive performance compared with normal birth weight controls at age 7 (Leijon, Ingemansson, Nelson, Wadsby, & Samuelsson, 2016). More recently, in a meta-analysis of 19 studies from South Asia, Upadhyay et al. (2019) reported that children under 10 years of age with LBW had significantly lower cognitive and motor development scores than normal birthweight controls.

Another key factor cognitive development is breastfeeding, both directly, via improved infant nutrition, and indirectly, through its impact on mother-infant bonding. A Lancet review and meta-analysis found that breastfeeding was consistently linked with better performance on intelligence tests in children and adolescents and reports a significant pooled increase in intelligence scores from 16 observational studies (Victora et al., 2016). In the Pelotas birth cohort in Brazil, results revealed a dose-response association between breastfeeding duration and a number of developmental indicators, including increased child intelligence, better educational outcomes, and higher income at the age of 30 years (Victora et al., 2015). The importance of breastfeeding has also been shown in the impact of interventions that encourage its uptake on cognitive development (Britto et al., 2017).

1.7: Summary of HIC literature review

In the broadest terms, this review has synthesised the existing evidence regarding the relationship between maternal depression throughout infancy and infant cognitive outcomes in HICs. More specifically, this review has examined whether this hypothesised effect is more dependent on early postnatal or chronic exposure to maternal depression. A timing effect was defined as the effect of single episode of depression occurring within the first postnatal year, while chronicity was defined as exposure to multiple or consistent

maternal depressive symptoms in the first two to three years of a child's life. The indirect and independent effects of other relevant factors were also discussed and reviewed.

There is some evidence for the effect of exposure to a single episode of depression in the postnatal period. Findings are inconsistent and many of the reviewed studies utilised small, low-risk samples but results do seem to indicate that exposure in the early postnatal period (< 3 months) may be more detrimental to development than later exposure. Overall, the weight of evidence points toward exposure to chronic depression throughout infancy as a more substantial risk to cognitive development. Maternal depression also appears to represent a greater risk for male infants than female infants, and findings from a number of large studies indicate that this risk is exerted through compromised maternal caregiving. While compromised caregiving can have a detrimental impact on the cognitive development of both male and female infants, there is evidence that male infants generally require a higher level of other-regulation and parental scaffolding to support their development that may be lacking in the context of maternal depression.

There is also some evidence that caregiving is only compromised by maternal depression in the context of low socioeconomic status. While high SES can serve to buffer against the risk posed by maternal depression, low SES may make both mothers and infants more vulnerable and represents the dual-risk of reducing the resources available to the mother and increasing the number of competing demands on her time. Finally, there is considerable evidence that maternal IQ and education, maternal age, and parity can have significant independent effects on infant cognitive development.

1.8 Focus of the current study

While studies from HICs are vital in terms of providing a framework that identifies variables and relationships of interest that are likely to be influencing infant development in LMICs, it cannot be assumed that risk and protective factors are operating in the same way in both contexts. As a result, good quality studies are required to firmly establish the scale of need and the important aetiological factors relevant to different LMIC settings. Now that a broad framework of expectations has been established from the HIC literature, the first part of this thesis will focus on an investigation of the impact of early maternal postnatal depression and chronic exposure through the first two years of life on infant cognitive development in India, and whether factors such as maternal caregiving and infant sex sit in the same roles identified in HIC studies. First, a systematic review of studies from LMIC settings will be presented to provide a more defined shape to expectations regarding the

roles of different risk and protective factors in these settings. Following this, findings from the new empirical study in India will be reported and discussed against the backdrop of existing the LMIC and HIC literature.

Specifically, the empirical study will examine the effect of a single episode of depression during the postnatal period and then the effect of chronic exposure over the first 2 years of an infant's life on cognitive outcomes at age 2. Contingent on the presence of a main effect of maternal depression, the mediating role of maternal sensitivity will also be investigated. Otherwise, the interaction between depression and sensitivity will be tested. Based in prior evidence of sex-specific effects, the study will explore these effects in the whole sample and then for boys and girls separately. Although the other variables discussed may also function as moderators of the association, it is beyond the scope of the current project to investigate such effects. Accordingly, antenatal depression, SES, maternal education, maternal age and parity will be treated as covariates. Concurrent depression will not be included as a covariate but will be included in the construction of a chronic depression factor. Furthermore, as the study will utilise independently rated observed cognitive outcomes, the risk of confounding from concurrent depression is significantly reduced.

The second part of the thesis will then focus on investigating how the Edinburgh Postnatal Depression Scale functions in an urban population in India, relative to an urban population in the UK. This will provide a better understanding of how mothers report postnatal depression in India and provide invaluable insights regarding how to compare and interpret findings from these very distinct cultural contexts. Findings will facilitate understanding at two levels, firstly in terms of direct comparisons of latent means, and secondly in terms of interpreting the relationships between different risk and protective factors in each setting.

Chapter 2: A systematic review of the association between perinatal depression and cognitive development in infancy in low- and middle-income countries[1].

2.1 Abstract

The association between perinatal depression and infant cognitive development has been well documented in research based in high-income contexts, but the literature regarding the same relationship in low and middle-income countries (LMICs) is less developed. The aim of this study is to systematically review what is known in this area in order to inform priorities for early intervention and future research in LMICs. The review protocol was pre-registered on Prospero (CRD42018108589) and relevant electronic databases were searched using a consistent set of keywords and 1473 articles were screened against the eligibility criteria. Sixteen articles were included in the review, seven focusing on the antenatal period, eight on the postnatal period, and one which included both. Five out of eight studies found a significant effect of antenatal depression (d = .21-.93) on infant cognitive development, while four out of nine studies found a significant effect of postnatal depression (d = .17-.47). Although the evidence suggests that LMICs should prioritise antenatal mental health care, many of the studies did not adequately isolate the effects of depression in each period. Furthermore, very few studies explored more complex interactions that may exist between perinatal depression and other relevant factors. More high-quality studies are needed in LMIC settings, driven by current theory, that test main effects and examine moderating or mediating pathways to cognitive development.

[1] Submitted for publication: Bluett-Duncan, M., Kishore, M.T., Patil, D. M., Satyanarayana, V.A., & Sharp, H. (pending review). A systematic review of the association between perinatal depression and cognitive development in infancy in low- and middle-income countries. *PLOS ONE*.

2.2 Introduction

The global prevalence of perinatal depression has been estimated at 11.9%, but a recent meta-regression has shown a significant difference between high income countries (HICs) and low- and middle-income countries (LMICs) (Woody, Ferrari, & Siskind, 2017). According to this review, the mean adjusted pooled prevalence of perinatal depression is 11.4% in HICs and 13.1% in LMICs. These findings are supported by Fisher et al. (2012) who also report that the prevalence of common perinatal mental disorders in LMICs is higher than that found in HICs, with weighted mean prevalence reported as 15.9% antenatally and 19.8% postnatally. Both studies drew data from studies that used a combination of diagnostic instruments and symptoms scales.

2.2.1 Sensitive period for development

Cognitive development refers to the whole range of mental activities and skills, including memory, language, learning, problem solving, perception and social cognition (Kihlstrom, 2018). During infancy cognition is typically assessed in terms of developmental milestones, including motor actions and language skills, and is conceptualised broadly as cognitive development, rather than the specific higher order functions, such as executive function and complex problem solving, that emerge later (Guerra, Williamson, & Lucas-Molina, 2012). In view of this, and in view of the developing nature of the literature in LMIC settings, this review utilises an inclusive and focused conceptualisation of global infant cognitive development that includes early mental, language and psychomotor milestones.

The perinatal period, and beyond into infancy, has received particular attention in HIC research because it represents an important and sensitive stage where children are at their most receptive stage of development (Black et al., 2017; Daelmans et al., 2017; Doyle, Harmon, Heckman, & Tremblay, 2009). From conception, up until around 3 years, the brain exhibits a high level of neural plasticity and rapid synapse formation that enhance the capacity of the child to learn and develop (Hannon, 2003). Neural connections are formed that affect development throughout the life course and provide the basis for future social, emotional and cognitive development (Sameroff, 2010). While this is beneficial for development under optimal conditions, adverse life experiences during this time can have long-lasting, detrimental effects (Teicher & Samson, 2016).

2.1.2 High-income settings

The possible detrimental effects of perinatal depression on infant health and development have been well documented in research from HICs. Significant deleterious impacts have

been observed in brain development, regulatory behaviours, acquisition of developmental milestones and various other behavioural, emotional, and physical outcomes (Glover, 2014; Hoffman, Dunn, & Njoroge, 2017; Wachs, Black, & Engle, 2009). In addition, there is a significant body of evidence supporting the presence of an association between perinatal depression and poor infant cognitive outcomes, summarised in a number of systematic reviews and meta-analyses (Beck, 1998; Grace, Evindar, & Stewart, 2003; Kingston, Tough, & Whitfield, 2012; Kingston, McDonald, Austin, & Touch, 2015; Liu et al., 2017; Van den Bergh et al., 2016). However, although generally significant, effects are generally small and long-term effects tend to be confined to high-risk samples and subgroups of infants experiencing additional risks, suggesting the presence of a complex interaction of different factors, rather than a simple and direct relationship (Graignic-Philippe, Dayan, Chokron, Jacquet, & Tordjman, 2014; Murray, Fearon, & Cooper, 2015). As such, developmental frameworks have become progressively more complex and HIC research has increasingly turned to the investigation of different mechanisms underlying the transmission of risk from perinatal depression to impaired infant cognitive development (Goodman & Gotlib, 1999; Sameroff, 2010). These mechanisms are not the focus of this review, but detailed discussion of the relevant processes can be found elsewhere (Murray, Fearon, & Cooper, 2015; Sameroff, 2010).

2.1.3 Low- and middle-income settings

In LMICs, where infants are generally exposed to a larger number and range of adversities, the impact of factors such as perinatal depression are likely to be magnified (Britto et al., 2017). Although depression may act in a similar way regardless of context, the co-occurrence of additional risk factors in LMICs is anticipated to result in increasingly compromised development, relative to that observed in HICs (Black et al., 2017; Walker et al., 2007). While a number of broader systematic reviews have included the relationship between perinatal depression and cognitive development in LMIC settings (Gelaye, Rondon, Araya, & Williams, 2016; Herba, Glover, Ramchandani, & Rondon, 2016; Parsons, Young, Rochat, Kringelbach, & Stein, 2012; Stein et al., 2014), there has not yet been a focused synthesis of studies on perinatal depression and cognitive development conducted in the LMIC context. A recent systematic review of studies examining the impact of perinatal mental health on infant neurodevelopment in LMIC settings (Burger, Hoosain, Unger, & Neihaus, 2020) included both a wide range of perinatal mental health disorders (e.g. schizophrenia, mania, PTSD, anxiety, depression) which may have very distinct effects on infant development, and a wide range of child outcomes to age 2 (gross and fine motor,

cognitive, language, behavior and social-emotional development). The narrative synthesis therefore did not provide an in-depth evaluation of the evidence for effects of perinatal depression per se on cognitive development specifically. The authors concluded "Due to heterogeneity of reported types of maternal health disorders and different domains of developmental outcomes, it was not possible to draw a definitive conclusion about the association between prenatal exposure to maternal mental health and child development" (Burger et al., 2020, p.168).

Finally, Although there now have been a number of individual studies that have explored the relationship between perinatal depression and infant cognitive development in LMIC settings, the overall literature in this context is not yet well developed and has not been synthesised. As such, it is expected that the majority of research identified in this review will be focused on investigating the presence of main effects of perinatal depression on infant outcomes, rather than on testing the more complex interplay between factors that has more recently emerged in the HIC literature.

2.1.4 Identifying independent effects of depression at different time-points

An important distinction demonstrated in research from HICs is that depression during the antenatal and postnatal periods exert independent effects on development and act on development through different mechanisms (Osborne et al., 2018). Accordingly, it is important to delineate independent and cumulative effects. In order to establish whether there is an independent effect within these distinct periods, studies need to satisfy two criteria. Firstly, because the presence of a depressive episode in one period is significantly associated with a depressive episode in the other, studies focusing on the influence of one period should adequately control for the effects of the other (Norhayati, Hazlina, Asrenee, & Emilin, 2015; Rahman & Creed, 2007). Secondly, studies investigating depression in either period on later cognitive outcome should account for the effects of concurrent depressive symptoms at the time of cognitive testing. Perinatal depression is known to increase the likelihood of future episodes of depression so any apparent cognitive impairments may be the result of current, rather than prior perinatal symptoms of depression, reflecting either a true impact on cognition, when assessed by independent observers, or, more often, a view of the child that is distorted by depressive symptoms when cognition is assessed via maternal report (Goodman, 2007; Howard et al., 2014; Richters & Pellegrini, 1989).

2.1.5 Aims of this review

The primary purpose of this review was to synthesise the evidence for the independent and joint effects of antenatal and postnatal depression on the cognitive development of infants aged 0-3 years in LMICs. The strength of the evidence for the independent effects will be discussed in terms of whether studies have adequately controlled for concurrent depressive symptoms and depressive episodes at other time-points in the perinatal period. These findings can then be used to inform recommendations for early intervention and the progression of the literature in this area in terms of how studies in LMICs can be designed to more effectively test current theory.

2.3 Materials and Methods

This research adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for systematic review (Moher et al., 2009). A protocol was completed prior to the review being carried out and registered on Prospero (CRD42018108589).

2.3.1 Inclusion and exclusion criteria

Observational studies (prospective, cross-sectional, case-control) were eligible for inclusion if they quantitatively assessed the association between maternal depression during pregnancy and/or the first 12 months post-birth and cognitive development in infants aged 0-3 years, and if a) maternal depression was diagnosed through a standardised diagnostic interview, or a validated symptom questionnaire or screening tool, b) cognitive development was assessed using a direct, validated measure of cognitive or language ability, and c) the research was carried out in a LMIC as designated by the World Bank.

Due to the emerging nature of the literature, this review has leaned toward overinclusion of studies, and therefore RCTs were also eligible for inclusion if they a) presented findings on the control arm of the study, b) present findings using baseline data, or c) they appropriately adjusted for intervention effects in the full sample. No historical time-limits were imposed on publication dates. Final searches included studies published in May 2020.

Studies were excluded if a) depressive symptoms were assessed as part of a general mental health assessment from which it was not possible to isolate the effects of depression (e.g. SRQ-20), b) they examined specific patient populations (mother or infant) with any physical disease or disorder that may impact cognitive development, other than maternal depression, c) development was assessed using implied measures of cognition, or d) they were case-series' or case-studies.

2.3.2 Definitions

Perinatal Depression: Defined as a non-psychotic depressive episode of mild to major severity that occurs during pregnancy or up to 12 months postnatal (Gelaye et al., 2016), diagnosed through standardised diagnostic interviews or validated symptom scales, and not restricted to confirmatory clinical diagnoses.

Infant Cognitive Development: This review utilises an inclusive conceptualisation of global infant cognitive development that includes early mental, language and psychomotor milestones.

Low- and Middle-Income Country (LMIC): A country designated as a LMIC by the World Bank at the point of data collection.

2.3.3 Search strategy and study selection

Eligible studies were identified using electronic and manual searches in October 2018. Follow-up searches were completed in May 2020 to ensure that all relevant studies were included. PubMed, PsychInfo and CINAHL were identified as the most relevant databases and searched for relevant articles. Keywords relating to the research questions were identified and an overall search strategy was devised using Boolean operators to combine free text and MeSH terms (see Figure 2.3.3 for example). Free text terms were used uniformly across searches, but minor alterations were made to tailor MeSH terms appropriately to each database. In order to reduce the impact of publication bias, the grey literature was also searched via ProQuest. Finally, reference lists of included articles were hand-searched to identify any relevant articles not revealed by the electronic search.

Study selection was completed in two phases. First, title and abstract screening of all retrieved articles was completed independently by two reviewers (MBD and DP). Studies which were deemed by both reviewers to meet inclusion criteria, or that were unclear, were selected for full-text review. Studies that clearly did not meet inclusion criteria were excluded at this point. Second, a full-text review of the remaining articles was completed independently by both reviewers. Studies that were deemed to meet the inclusion criteria at this point were included in the final review. All final decisions were made by consensus between the two reviewers. Where studies remained unclear or there were disagreements between the two reviewers, articles were referred to a third reviewer (HS or TK) who independently reviewed the articles and participated in the group consensus. Reasons for exclusion were recorded.

Figure 2.3.3: Example search strategy used with PubMed[2]

Search #1	((Mothers [Text Word] OR Maternal [Text Word] OR Perinatal [Text Word] OR Peripartum [Text Word] OR Prenatal [Text Word] OR Antenatal [Text Word] OR Antepartum [Text Word] OR Pregnancy [Text Word] OR Pregnant [Text Word] OR Trimester [Text Word] OR Postnatal [Text Word] OR Postpartum [Text Word] OR Puerperal [Text Word] OR Puerperium [Text Word] OR post-birth [Text Word])) OR (Peripartum [mh] OR Postpartum period [mh:noexp] OR Prenatal care [mh] OR Postnatal care [mh] OR Postpartum depression [mh])
Search #2	Depression [Text Word] OR Depressed [Text Word] OR depressive symptoms [Text Word] OR Depression symptoms [Text Word] OR affective disorder [Text Word] OR depressive disorder [Text Word] OR depression [mh] OR postpartum depression [mh] OR depressive disorder [mh]
Search #3	1 AND 2
Search #4	Infant [Text Word] OR Infants [Text Word] OR child [Text Word] OR children [Text Word] OR infancy [Text Word] OR baby [Text Word] OR babies [Text Word] OR toddler [Text Word] OR Toddlers [Text Word] OR newborn [Text Word] OR childhood [Text Word] OR pre-school [Text Word] OR infant [mh] OR infant, newborn [mh:noexp] OR child, preschool [mh]
Search #5	Cognition [Text Word] OR cognitive [Text Word] OR language [Text Word] OR IQ [Text Word] OR intelligence [Text Word] OR memory [Text Word] OR perception [Text Word] OR learning [Text Word] OR problem solving [Text Word] OR metacognition [Text Word] OR social cognition [Text Word] OR DQ [Text Word] OR communication [Text Word] OR executive function [Text Word] OR attention [Text Word] OR child development [mh] OR cognition [mh] OR intelligence [mh]
Search #6	4 AND 5
Search #7	3 AND 6
Search #8	low income population [Text Word] OR low income countr* [Text Word] OR middle income population [Text Word] OR middle income countr* [Text Word] OR low and middle income population [Text Word] OR low and middle income countr* [Text Word] OR developing countr* [Text Word] OR developing nations [Text Word] OR third world [Text Word] poverty [Text Word] OR LMIC [Text Word] OR LAMIC [Text Word] OR africa [Text Word] OR asia [Text Word] OR south america [Text Word] OR central america [Text Word] OR South Asia [Text Word] OR middle east [Text Word] OR poverty [mh] OR Asia [mh] OR Africa [mh] OR south America [mh] OR central America [mh]
Search #9	All countries designated as LMIC by World Bank included here as individual Text Words.
Search #10	8 OR 9
Search #11	7 AND 10

2.3.4 Quality assessment

Study quality was evaluated independently by two reviewers (MBD and DP) using the Newcastle-Ottawa Scale (NOS) adapted for cohort studies (Wells et al., 2013). Studies were appraised in 3 areas: selection, comparability, and outcome. Each of these had a number of

[2] Note: For full details of search strategy for each database see Appendix 3

criteria which were scored according to NOS guidelines, producing quality ratings of good, fair and poor. Each study was assessed using a standardised form which included instructions and clarifications where necessary for applying the scale to the current subject. One of the criteria from the "selection" section (*"Demonstration that outcome of interest was not present at the start of the study")* was not appropriate to the study design in question since many aspects of cognitive function cannot be assessed early in the perinatal period. However, rather than remove this item and adjust the standardised scoring system, it was decided to simply give all the studies a score of zero for this item. Thus, it may be that the quality assessments slightly underestimate the quality of each study. Following independent review, any disagreements regarding quality were solved by consensus.

2.3.5 Data extraction

Data extraction was completed independently by two reviewers (MBD and DP) using a standardised data extraction form that was created specifically for this review. Data extracted included country/setting, study design, aims/hypotheses, sample size, sample characteristics, measurement of exposure, measurement of outcome, perinatal stage at exposure, infant age at outcome, statistical approach, confounding variables, mediating or moderating variables, and main results (descriptive, unadjusted, adjusted). Following an initial piloting phase, data were extracted and compared. Any disagreements were resolved by consensus, and any issues that still lacked clarity were referred to HS. Where relevant data were missing study authors were contacted to request specific information.

2.3.6 Analysis

Extraction included the recording of variables that were defined a priori: socioeconomic status, maternal age, maternal education, infant gender, birthweight, intervention (if RCT). If meta-analysis had been possible this set would have represented a minimum set of adjustment variables to compare the effect size of "adequately" controlled studies and "inadequately" controlled studies. In practice, meta-analysis was not possible due to the variation between studies in terms of their context and the measurement tools used. However, this set of variables were considered when examining patterns in the data through narrative synthesis. Articles presenting findings regarding exposure to antenatal and postnatal depression were synthesised and discussed separately.

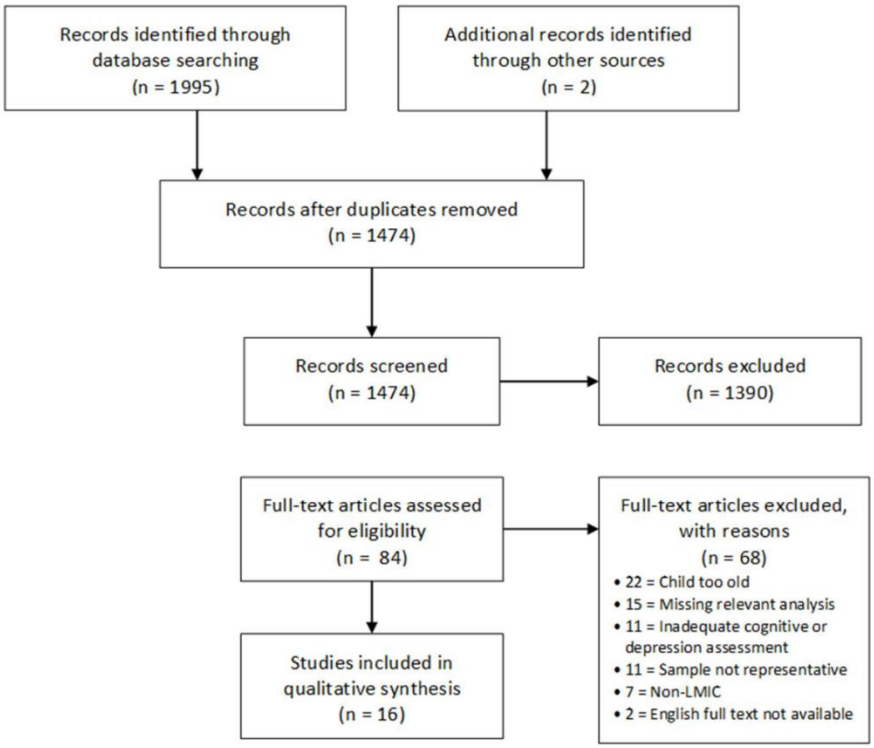
The adjusted standardised mean difference (d) between depressed and non-depressed group was calculated wherever possible. If this was not possible, unadjusted effect sizes

were calculated using group means and standard deviations. The strength of the effect size can be interpreted as follows: 0.2 = small; 0.5 = medium; 0.8 = large (Lakens, 2013).

2.4. Results

16 studies were selected for inclusion in this review (Figure 2.4), 7 with a primary focus on antenatal depression, 8 with a primary focus on postnatal depression, and 1 study that considered both pre- and postnatal exposure. The majority of studies utilised a prospective-cohort design (10). A number of Randomised Controlled Trials (6) were also included if they reported on data from the control arm of the study or they controlled for the effect of the intervention in the final analysis. Finally, there was one study which used a quasi-experimental design.

Figure 2.4: Study flow diagram



2.4.1 Approach to narrative synthesis

Findings are grouped by perinatal period according to the main focus of each paper. Significant results are presented first, followed by non-significant findings. Findings are then briefly synthesised in terms of the adherence of each study to the requirements for assessing the independent effects of maternal depression within each period and the inclusion by each study of the a priori key adjustment variables.

Unless otherwise stated, beta coefficients given for main effects are unstandardized and refer to the change in the number of points scored on a particular measure when a mother is depressed, compared to when a mother is non-depressed. Means and standard deviations are reported where available. All depression assessments are maternal self-report unless otherwise stated.

2.4.2 Summary of antenatal studies

Table 2.4.2 provides a summary of the key features of the 8 antenatal studies. Antenatal studies were conducted in five different countries: four in South Africa and one in each of China, Mexico, Ukraine and Vietnam. Exposure to maternal depressive symptoms was assessed in four studies by the Edinburgh Postnatal Depression Scale (EPDS) (Cox, Holden, & Sagovsky, 1987), twice with the Beck Depression Inventory (BDI-I and BDI-II) (Beck, et al., 1961, 1996), once with the Symptom Checklist-90-Revised (SC-90-R) (Derogatis & Savitz, 1999) and once with the Structured Clinical Interview for DSM-IV diagnoses (SCID) (First & Gibbon, 2004). Six studies specified that depression was assessed during the 3rd trimester, one during the 2nd trimester and one simply stated that the assessment was conducted during the antenatal period. The Bayley Scales of Infant Development (BSID) were most commonly utilised for assessing infant outcome, with four studies using the BSID-III (Bayley, 2006) and two studies using the BSID-II (Bayley, 1993). Another study used the Gesell Scale (Williamson, Wilson, Lifschitz, & Thurber, 1990) and the remaining study used the Peabody Picture Vocabulary Test (PPVT) (Dunn, 1965) and an executive function battery (EF) (Blair, Zelazo, & Greenberg, 2005). Age at follow-up ranged from 6-30 months. One study was rated as good, six as fair, and one as poor quality, according to the NOS.

2.4.2.1 Significant findings

Five out of eight studies found a significant effect of antenatal depression on infant cognitive development (see Appendix 3 for full summary of results). Four of these studies were fair quality (Donald et al., 2019; Lin et al., 2017; Munoz-Rocha et al., 2018; Tran et al., 2013) and one was poor quality (Breen et al., 2018).

Arguably the most methodologically robust study in this review is Tran et al. (2013). Although the study only received a quality rating of fair, it was the only study to control for both postnatal and concurrent depression. Using a prospective design to assess a sample in Vietnam, this study found a small effect of depression in that infants of depressed mothers (M = 97.9, S.D. = 14.1) scored lower than infants of non-depressed mothers (M = 100.0, S.D. = 12.8, *d* = 0.15) on the BSD-III. Importantly, the authors demonstrated that a significant

Table 2.4.2:

Key features of antenatal studies

Authors (Year)	Country	Design	Sample (follow up %)	Exposure (Cut-off)	Exposure Timing	Outcome	Outcome Subscales	Outcome Timing	Study Quality
Bandoli et al., (2016)	Ukraine	RCT	754 (45.6)	BDI (≥10)	32 weeks gestation	BSID-II	MDI	6 & 12 months	Fair
Breen et al., (2018)	South Africa	Prospective	149 (77.2)	BDI-II (≥15)	28-32 weeks gestation	BSID-III	Cognitive Language	24 months	Poor
Donald et al., (2019)	South Africa	Prospective	1143 (73)	EPDS (>13)	28-32 weeks gestation	BSID-III	Cognitive Language	24 months	Fair
Lin et al., (2017)	China	Prospective	398 (56.5)	SCL-90-R (N/A)	28-36 weeks gestation	Gesell Scale	Language	24-30 months	Fair
Munoz-Rocha et al., (2018)	Mexico	Prospective	760 (62.2)	EPDS (≥13)	3rd Trimester	BSID-III	Cognitive Language	24-30 months	Fair
Murray et al., (2016)	South Africa	RCT	449 (59.0)	SCID (N/A)	Antenatal period	BSID-II	MDI	18 months	Good
Rotheram-Fuller et al., (2018)	South Africa	RCT	1238 (90.0)	EPDS (≥13)	26 weeks gestation.	Executive Function Battery PPVT	OS SS STS N/A	36 months 36 months	Fair
Tran et al., (2013)	Vietnam	Prospective	497 (76.1)	EPDS-Vietnam (≥4)	>28 weeks gestation.	BSID-III	Cognitive	6 months	Fair

Key: BDI = Beck Depression Inventory; EPDS = Edinburgh Postnatal Depression Scale; SCL = Symptom Checklist; SCID = Structured Clinical Interview for DSM-IV; BSID = Bayley Scales of Infant Development; PPVT = Peabody Picture Vocabulary Test; MDI = Mental Development Index; OS = Operation Span; SS = Silly Sounds; STS = Something's the Same.

association between depression and cognitive development at 6 months remained after controlling for both postpartum and concurrent depressive symptoms (B = -4.80, p < 0.05). Lin et al. (2017) also controlled for concurrent depression but not postnatal depression. The authors assessed a relatively high SES sample in China using a prospective design. Depression was associated with significant deficits in language scores on the Gesell Scale at 24-30 months (B = -13.18, p = 0.01). Unfortunately, no cognitive composite score was reported which means that the findings are not easily comparable with other studies.

The three other studies that found a significant effect of antenatal depression on cognitive outcomes did not account for postnatal or concurrent depression in their analyses and so their results need to be interpreted with caution. Outside of this consideration, however, all are robust studies and are discussed next.

Two studies utilised data from the same prospective sample, the Drakenstein Health Study in South Africa. Breen *et al.* (2018) used a nested sub-sample while Donald et al. (2019) used data from the whole sample. Interestingly, the two studies used different tools to assess antenatal depression, with Breen et al. using the BDI-II and Donald et al. using the EPDS. Breen et al. found that, at 24 months, infants of mothers who were depressed scored significantly lower on the cognitive subscale of the BSID-III (M = 83.7, S.D. = 6.2) than infants of non-depressed mothers (M = 90.0, S.D = 7.8, *d* = 0.93) and they also scored significantly lower on the language subscale (M = 83.8, S.D. = 14.1) than those of the non-depressed group (M = 87.1, S.D. = 9.9, *d* = 0.29). In the full Drakenstein Health Study sample, Donald et al. reported the unadjusted and adjusted effects of antenatal depression on cognitive and language outcomes, also using the BSID-III. Unadjusted bivariate linear regression analyses were carried out on the raw scores of the cognitive, receptive language and expressive language subscales of the BSID-III for the full sample and after splitting the sample by gender. They reported a significant effect of antenatal depression on cognitive outcomes in boys only (B = -1.58, p < 0.05), and on expressive language for the full sample (B = -1.06, p < 0.05) and girls only (B = -1.96, p < 0.05). Adjusted analyses revealed a significant effect of antenatal depression on cognitive outcomes at 24 months for the full sample after adjusting for a comprehensive set of confounders (β = −1.03, p = 0.027, *d* = 0.21). Final results for the language subscale were not reported.

Similarly, Munoz-Rocha et al. (2018) found evidence for a significant effect of antenatal depression on both cognitive and language development at 24 months in a Mexican cohort study. This study was designed to investigate the role of blood manganese at different

points in the development of the foetus and incorporated this into their final models. In the first model, which included 3rd trimester blood manganese as a covariate, depression during the 3rd trimester was associated with significant deficits in both cognitive (B = -2.40, $p < 0.01$) and language skills (B = -2.47, $p < 0.01$) assessed by the BSID-III. However, in the second model, which excluded antenatal levels of blood manganese but included levels of cord blood manganese during birth, the coefficient was of similar magnitude but was only marginally significant for both cognitive (B = -2.20, $p = 0.06$) and language skills (B = -2.17, $p = 0.08$).

2.4.2.2 Non-significant findings

The three studies that did not find a significant effect of antenatal depression received good (Murray, Cooper, Arteche, Stein, & Tomlinson, 2016) and fair (Bandoli et al., 2016; Rotheram-Fuller et al., 2018) quality ratings.

The key difference between Murray et al. (2016) and many of the others in this review is that exposure to antenatal depression was confirmed via diagnostic interview (SCID). In an RCT from South Africa the authors first report an unadjusted ANOVA showing a significant difference between BISID-II mental development index scores for infants of depressed (M = 81.4, S.D. = 10.1) and non-depressed mothers (M = 84.8, S.D. = 5.1) ($F(1,262) = 4.4$, $p = 0.04$, $d = 0.33$). This moderate effect remained significant after adjusting for 2 month ($p < 0.03$) and 6 month ($p < 0.05$) postnatal depression. However, the effect size was attenuated and became marginal once intervention, risk and the intervention by risk interaction term were included in the final model ($F = 3.1$, $p = 0.08$, $d = 0.27$). Risk was a composite measure that included items relating to maternal age, maternal education and SES. Although non-significant, these findings are indicative of a role for antenatal depression in infant cognitive development.

Rotheram-Fuller et al. (2018) utilised data from an RCT in South Africa that was investigating the efficacy of a home visiting intervention carried out by community health workers. The sample was split into four groups relating to the period during which depression was present: never depressed, antenatal, postnatal, antenatal/postnatal. Although this approach does help to isolate the effects of antenatal depression from exposure to postnatal and chronic depression, the postnatal depression group was created from assessments at 2 weeks and 6, 18 and 36 months, which goes beyond the cut-off of 12 months defined as the postnatal period in the inclusion criteria for this review. There were no significant differences between groups at 36 months on the PPVT or on two of the

EF battery tasks. This study did not control for the intervention but have reported that there was no effect of the intervention, independently or in combination with maternal depressive group, on cognitive outcome. Bonferroni's method for computing p values and confidence intervals was used to account for multiple comparisons.

Finally, Bandoli et al. (2016) found no evidence of an effect of antenatal depression in an RCT based in Ukraine. It is important to note that the authors were also interested in the effect of periconceptional alcohol use and so the sample included a higher proportion of moderate to heavy drinkers than would normally be found in a general population. Subsequently, final models included periconceptional alcohol use as a covariate. The authors reported a non-significant effect of exposure to depressive symptoms at 32 weeks gestation on BSID-II mental development index scores at 6 months (B =-1.96, $p > 0.05$) and at 12 months (B = -0.16, $p > 0.05$). There was also no significant mean difference between groups when restricted to mothers who had not used alcohol at or around conception at either 6 or 12 months.

2.4.3 Summary of postnatal studies

Table 2.4.3 provides a summary of the key features of the postnatal studies. Postnatal studies were conducted in 8 different countries: two in Bangladesh, and one in each of Barbados, Brazil, India, Pakistan, South Africa, Uganda and Vietnam. Exposure to maternal depressive symptoms was assessed using the EPDS in three studies, the Centre for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977) in two studies, and the Zung Depression and Anxiety Scale (ZDAS) (Zung, 1965, 1971), the Aga Khan University Anxiety and Depression Scale (AKUADS) (Ali, Khan, & Jehan, 1998) and the Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998) in one study each. Infant outcome was assessed with a wide range of tools. The BSID-II was utilised in three studies and the BSID-III in one study. The Mullen Scales of Early Learning (MSEL) (Mullen, 1995), the Griffiths Mental Development Scale (Griffiths, 1954), the Early Childhood Development Tool (ECD), and the Developmental Assessment Scales for Indian Infants (DASII) (Patni, 2012), and the executive function battery were all used once. Exposure to depression was assessed from 2 months to 12 months and age at follow up ranged for 3 months to 36 months. Several studies assessed the child at more than one time-point or with more than one tool.

Table 2.4.3:

Key features of postnatal studies

Authors (year)	Country	Design	Sample (follow-up %)	Exposure (cut-off)	Exposure Timing	Outcome	Outcome Subscale	Outcome Timing	Study Quality
Ali et al., (2013)	Pakistan	Quasi-Experimental	420 (91.6)	AKUADS (≥17)	1, 2, 6, 12, 18, 24 & 36m	ECD Tool	Cognitive & Language	1-12m (monthly), 18, 24, 30 & 36m	Poor/Good
Black et al., (2007)	Bangladesh	RCT	346 (63.9)	CES-D (N/A)	12m	BSID-II	MDI	6 & 12m	Fair
Familiar et al., (2018)	Uganda	Prospective	228 (95.6)	CES-D (≥16)	6m	MSEL	Cognitive Composite	6 & 12m	Poor
Galler et al., (2000)	Barbados	Prospective	225 (48.5)	ZDAS (≥50)	7w & 6m	Griffiths Mental Development Scale	Developmental Quotient	7w, 3 & 6m	Fair
Garman et al., (2019)	South Africa	RCT-Control	594 (62.8)	EPDS (N/A)	2w, 6 & 18m	BSID-II	MDI	18m	Fair
						Executive Function Battery	OS, SS, STS	36m	
Hamadani et al., (2012)	Bangladesh	Prospective	488 (100.0)	EPDS (≥10)	6w & 6m	BSID-II	MDI	12m	Fair
Patel et al., (2003)	India	Prospective	171 (52.0)	EPDS (≥11)	6w	DASII	MDI	6m	Fair
Quevedo et al., (2012)	Brazil	Prospective	342 (86.5)	MINI (N/A)	1-2m	BSID-III	Language	12m	Good
Tran et al., (2013)	Vietnam	Prospective	497 (76.1)	EPDS-Vietnam (≥4)	8w & 6m	BSID-III	Cognitive	6m	Fair

Key: AKUADS = Aga Khan University Anxiety and Depression Scale; CES-D = Centre for Epidemiological Studies – Depression; ZDAS = Zung Depression and Anxiety Scales; EPDS = Edinburgh Postnatal Depression Scale; MINI = Mini International Neuropsychiatric Interview; ECD Tool = Early Childhood Development Tool; BSID = Bayley Scale of Infant Development; MSEL = Mullen Scales of Early Learning; DASII = Developmental Assessment Scale for Indian Infants; MDI = Mental Development Index; OS = Operation Span; SS = Silly Sounds; STS = Something's the Same.

2.4.3.1 Significant findings

Out of nine studies reviewed, four found a main effect of postnatal depression (see Appendix 3 for full summary of results). One was good quality (Quevedo et al., 2012), two were fair quality (Galler, Harrison, Ramsey, Forde, & Butler, 2000; Patel, DeSouza, & Rodrigues, 2003) and one was poor quality (Ali, Mahmud, Khan, & Ali, 2013).

Quevedo et al. (2012) was the only postnatal study to receive a 'good' quality rating. Two characteristics set this study apart from many of the others. First, the authors took concurrent depression into account, and secondly, the authors used a diagnostic interview to assess maternal depression. Infants were assessed for language development at 12 months using the BSID-III and split into 4 groups according to mothers' depression. Group means and unadjusted effect size, with the no depression group used as the reference group, are reported in parentheses. The four groups were as follows, none (M = 108.6, S.D. = 17.0), postpartum (M = 107.2, S.D. = 16.5, *d* = 0.07), current (M = 105.9, S.D = 14.1, *d* = 0.16) and postpartum/current (97.4, S.D. = 15.4, *d* = 0.66). Adjusted analyses showed that infants of mothers who reported depression had significantly lower language scores (B = -2.87 $p < 0.01$, *d* = 0.17) and post-hoc tests revealed that this effect was only significant for the combined postpartum and current depression group.

All other studies with significant findings did not account for either antenatal or concurrent depressive symptoms. Patel et al. (2003) used a case-control design nested within a prospective cohort to assess mother-infant dyads in India. When DASII scores were treated as continuous, there was a significant difference between the mean cognitive scores of the depressed (M = 86.4, S.D = 8.33) and the non-depressed (M = 90.3, S.D. = 9.27, *d* = 0.44) groups. When treated as dichotomous, with scores <85 considered to be at risk for developmental delay, infants of postnatally depressed mothers were significantly more likely to be delayed than infants of non-depressed mothers (OR = 3.3, $p = 0.02$). Galler et al. (2000) assessed cognitive outcomes using the Griffiths Mental Development Scale in a prospective cohort from Barbados. Following a repeated measures analysis of cognitive outcomes at 7 weeks, 3 months and 6 months, the authors report a significant effect of depression at 7 weeks on cognitive outcomes at 3 months, after adjusting for confounders ($F(3, 78) = 2.09$; $p < 0.02$). Further, although not statistically significant, a similar linear relationship was also present between 7-week depression and 6 months cognitive outcomes.

Finally, Ali *et al.* (2013) utilised a quasi-experimental approach to assess developmental outcomes in Pakistan. Depression was assessed at 1, 2, 6, 12, 18, 24 and 36 months, while cognitive and language outcomes were assessed using the ECD every month in the first postnatal year and then at 18, 24, 30 and 36 months. The authors first looked for an interaction between PND and infant age on development and followed up on any significant interactions with a cross-sectional analysis of the impact of PND at 2, 6, and 12 months. Infants of depressed mothers were significantly more likely to be delayed at the 6th month of follow-up (OR= 3.3 (95% CI: 1.1, 9.9)) and the 12th month of follow-up (OR= 6.8 (95% CI: 3.0, 15.7)). There was no significant main effect of depression on language.

2.4.3.2 Non-significant findings

Five out of nine postnatal studies did not find a significant main effect of postnatal depression on cognitive outcomes. Of these five studies, four were fair quality (Black et al., 2007; Garman, Cols, Tomlinson, Rotheram-Borus, & Lund, 2019; Hamadani et al., 2012; Tran *et al*., 2013) and one was poor quality (Familiar et al., 2018).

Tran et al. (2013) was the only postnatal study to control for the effects of both antenatal and concurrent depressive symptoms. However, unlike the antenatal findings, the authors did not find a significant effect of postnatal depression on BSID-III cognitive scores at 6 months (B = 1.26, p > 0.05). Interestingly, the unadjusted means indicated that infants of depressed mothers (M = 102.9, S.D. = 14.1) actually scored higher than infants of non-depressed mothers (M = 98.5, S.D. = 13.1, *d* = 0.32).

Hamadani et al. (2012) found no significant difference in BSID-II mental development index scores between infants of mothers depressed at 6 weeks and/or 6 months (M = 99.7, S.D. = 10.8) and those reporting no depressive symptoms (M = 100.6, S.D = 12.1, *d* = 0.08) in a sample from Bangladesh. Familiar et al. (2018) assessed cognitive outcomes at 6 and 12 months in a Ugandan sample. Using a repeated measures approach, the authors found no main effect of depression on the composite cognitive score (B = -2.39, *p* = 0.15).

Garman et al. (2019) utilised data from the same South African RCT as Rotheram-Fuller et al. (2018) but drew data from the control arm of the study. Latent growth analysis was used to assess the effects of distinct trajectories of maternal depression. Mothers' depression was characterised as chronic low, early postpartum (6m), late postpartum (18m) and chronic high (6 & 18m). The authors report that there was no significant effect of early postpartum depression (M = 10.2, S.D. = 2.3) compared to the chronic low group (M = 10.1, S.D = 3.0, *d* = 0.08) on BSID-II mental development index scores at 18 months (β = 0.08 p =

0.905). There were also no significant effects on any of the tasks that formed the EF battery at 36 months (B's = -0.75, -0.29, -0.11, all *p*'s > 0.05). Similarly, there were no significant effects for any of the other depression groups. It is unclear whether beta coefficients reported are standardised or not.

Finally, Black et al. (2007) used a cross-sectional design to assess mother-infant dyads in Bangladesh who were participating in an RCT. Depressive symptoms and cognitive outcomes were assessed at 12 months using the CES-D and the BSID-II, respectively. The unadjusted main effect of depression was significant (r = -0.14, $p < 0.05$), but the adjusted model, including earlier cognitive development at 6 months, revealed a non-significant main effect (B = 0.09, p > 0.05). Although an unadjusted main effect is reported, this may reflect the effects of the micronutrients given within the RCT.

2.4.4 Key adjustment variables in postnatal and antenatal studies

As already described, only Tran et al. (2013) fully isolated the effects of depression within each period by adequately controlling for depression at other relevant time points. Four other studies at least partially controlled for either antenatal/postnatal or concurrent depression (Lin et al., 2017; Murray et al., 2016; Quevedo et al., 2012; Rotheram-Fuller et al., 2018). Overall, the majority of studies demonstrated low adherence to the requirements for isolating the effects of a particular episode of depression. This severely undermines confidence in the reported findings, particularly with regards to effects of postnatal exposure.

A number of additional "adjustment" variables were identified a priori in order to rule out third variable effects with relative confidence, including socioeconomic status, maternal age, maternal education, infant gender, birthweight, and intervention (if RCT). Tables 2.4.4a and 2.4.4b show which of these variables each study controlled for in their analyses. Only Donald et al. (2019) controlled for all of these variables, while Tran et al. (2013) controlled for all except infant gender. These two studies stand out as two of the more robust included in this review and are the least likely to be presenting results that could be confounded by a third variable. All other studies except one (Hamadani et al., 2012) included at least two in adjusted analyses. There were no clear patterns that emerged in the data with regards to the variables listed above.

Table 2.4.4a

Key *a priori* adjustment variables identified for antenatal studies.

Study	PND	Concurrent Depression	SES	Maternal Age	Maternal Education	Infant Gender	Birthweight
Bandoli et al., (2016)	No	No	**Yes**	**Yes**	**Yes**	No	No
Breen et al., (2018)	No	No	No	No	No	No	No
Donald et al., (2019_	No	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Lin et al. (2017)	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No
Munoz-Rocha et al., (2018)	No	No	No	**Yes**	**Yes**	**Yes**	**Yes**
Murray et al., (2016)	**Yes**	No	**Yes**	**Yes**	**Yes**	No	No
Rotheram-Fuller et al., (2018)	No	No	**Yes**	No	**Yes**	No	No
Tran et al., (2013)	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No	**Yes**

Table 2.4.4b

Key *a priori* adjustment variables identified for postnatal studies.

Study	AND	Concurrent	SES	Maternal Age	Maternal Education	Infant Gender	Birthweight
Ali et al., (2013)	No	N/A	No	**Yes**	**Yes**	No	No
Black et al., (2007)	No	N/A	**Yes**	**Yes**	**Yes**	**Yes**	No
Familiar et al., (2018)	No	No	**Yes**	**Yes**	**Yes**	**Yes**	No
Galler et al., (2000)	No	No	**Yes**	**Yes**	**Yes**	No	**Yes**
Garman et al., (2019)	No	No	**Yes**	**Yes**	**Yes**	No	No
Hamadani et al., (2012)	No	No	No	No	No	No	No
Patel et al., (2003)	No	No	No	No	**Yes**	No	**Yes**
Quevedo et al., (2012)	No	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**
Tran et al., (2013)	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No	**Yes**

2.5. Discussion

This systematic review of 16 studies investigating the association between maternal perinatal depression and infant cognitive outcomes in LMICs provides relatively strong evidence for an effect of antenatal depression but more inconsistent findings amongst postnatal studies. Five out of eight antenatal studies found a prospective negative main effect of depression on cognitive outcomes. However, only four out of nine postnatal studies found evidence of a prospective negative main effect on, or cross-sectional association with, infant cognitive development. While limited in number and mixed in design, most studies were of fair to good quality, with only three studies being rated as poor. It is important to consider and interpret this pattern of findings in the context of key study design features, as well as the size and quality of the evidence base, and the potential moderated effects that were not tested. This will be discussed in the following sections.

2.5.1 Is the perinatal period a sensitive period for cognitive development?

2.5.1.1 Antenatal depression

There is fairly consistent evidence of a significant effect of antenatal depression on infant cognitive development. All five antenatal studies that used a prospective, observational design report a significant effect (*d* = 0.21 – 0.93) of antenatal depression on infant cognitive outcomes (Breen et al., 2018; Donald et al., 2019; Lin et al., 2017; Munoz-Rocha et al., 2018; Tran et al., 2013). A quality rating of fair was given to four of these studies, and poor to one. Although, these findings echo the conclusions of Van Den Bergh et al. (2016) that exposure to various types of stress in utero, including depression, can have lasting effects on offspring development, the studies reviewed here actually find a more consistent effect of antenatal depression than studies in HICs, where there is a broadly even split between those finding a significant effect on cognitive development (Bhang et al., 2016; Deave, Heron, Evans, & Emond, 2008; Koutra et al., 2013; Evans et al., 2012; Schechter et al., 2017; Skurtveit, Selmer, Roth, Hernandez-Diaz, & Handal, 2014) and those that do not (Husain, Cruickshank, Tomenson, Khan, & Rahman, 2012; Ibanez et al., 2015; O'Leary et al., 2019; Otake et al., 2014; Tse et al., 2010). It is possible that these differences arise because antenatal depression is experienced alongside other risk factors more prevalent in LMIC settings, such as intimate partner violence or poor nutrition, that may exacerbate foetal programming effects (Herba et al., 2016).

Interestingly, the evidence suggests that the effects of antenatal depression are relatively long-term. Four antenatal studies found an effect of depression during the 3rd trimester on

cognitive development at 24-30 months. This is an important finding because it is between the 2nd and 3rd year that a child's cognitive ability begins to stabilise and form the foundation for future development (Doyle et al., 2009; Feldman & Eidelman, 2009). Thus, any impairment at this age is likely to have a lasting effect on child outcomes. This is discussed further in section 4.4.

Concerning the criteria for determining evidence of an independent effect of antenatal depression, adherence was generally low in the reviewed studies. One study controlled for postnatal depression only (Murray et al., 2016), one controlled for concurrent depression only (Lin et al, 2017), and one controlled for both postnatal and concurrent depressive symptoms (Tran et al., 2013). One additional study isolated the effects of antenatal depression by splitting the sample into groups based on the timing of exposure (Rotheram-Fuller et al., 2018). Although representing only a small proportion of the reviewed papers, it is worth noting that in all but one of these studies, the effect of antenatal depression remained significant even after controlling for postnatal or concurrent depression (although the main effect found in Murray et al. (2016) became marginal after including additional factors in the model).

Of particular note are the findings of Tran et al. (2013) who used path analysis to simultaneously test the effects antenatal, postnatal and concurrent depression, and report that only antenatal depression was a significant predictor of cognitive development. This suggests that the impact of antenatal depression on cognitive outcomes cannot be accounted for by postnatal or concurrent depressive symptoms. This study also controlled for other key factors such as maternal education, SES and infant birthweight, all of which have been identified as important factors in a child's cognitive development (Donald et al., 2019). Another key study is Donald et al. (2019). Although they did not control for postnatal or concurrent depression, they did control for each of the key adjustment variables that were identified a priori. By doing this, the authors have ruled out confounding effects from relevant third variables, thereby providing a good degree of confidence that the reported effect was in fact due to antenatal depression. It is important that these findings are replicated in equally well-designed studies.

All of the studies which did not find a significant main effect of antenatal depression were RCTs and this may have influenced results. In fact, this is clearly the case with Murray et al. (2016), where the initial significant effect of antenatal depression was attenuated ($p = .06$) after the inclusion of a risk by intervention interaction term in the model. Although neither

predicted cognition independently, exploration of the interaction between the two factors showed that the intervention improved cognitive performance in low-risk infants. This suggests that merely controlling for the effects of the intervention in an RCT may not adequately account for more complex interactive effects with other factors.

In the case of Bandoli et al. (2016), it is pertinent to consider the nature of the sample, which consisted of an index group who reported moderate to heavy drinking during the periconceptional period, and a control group with low or no exposure to alcohol at that point. It is possible that this influenced the results of the study, although there was no mean difference in cognitive scores for infants of depressed or non-depressed mothers within the control group either. Another consideration is that this study took place in Ukraine. Although categorised as a LMIC by the World Bank, Eastern European culture is very different from African or Asian culture, where most of the studies in review are drawn from, making direct comparisons with other studies more difficult.

Finally, while Rotheram-Fuller et al. (2018) were able to isolate the effects of antenatal depression, they did not actually control for the effects of the intervention. Although the authors investigated the effects of the intervention on cognitive development and found none, either independently or in interaction with maternal depressive status, it is possible that the intervention may have had some undetected influence on results.

2.5.1.2 Postnatal depression

Evidence for an independent postnatal effect is far less consistent. Three studies, all rated as fair quality, found a significant and negative prospective effect (*d* = .17 - .47; OR = 3.3 – 6.8) of postnatal depression (Galler et al., 2000; Patel et al., 2003; Quevedo et al., 2012), and one study, rated as poor, found a cross-sectional association (Ali et al., 2013). The remaining five postnatal studies, rated poor to fair quality, failed to find any direct association between maternal depressive symptoms and cognitive development, although three of these were RCT designs.

Tran et al. (2013) was the only study to control for antenatal and concurrent depressive symptoms and they did not find a significant effect of postnatal depression. Another study found that postnatal symptoms at 2 months and concurrent symptoms at 12 months did not independently predict language outcomes, but that there was an effect of chronic exposure to 2-month and 12-month depression (Quevedo et al., 2012). This finding illustrates the importance of investigating the effects of concurrent depressive symptoms and corresponds to findings from HIC literature that concurrent depressive symptoms may

account for the perceived effect of postnatal depression on cognitive outcomes (Sutter-Dalley et al., 2011).

Although a recent meta-analysis found a significant but small effect of postnatal depression in HICs (Liu et al., 2017), closer inspection of individual studies suggests that the effect is generally more apparent and pronounced in high-risk samples (Conroy et al., 2012; Hay et al., 2001; Lyons-Ruth, Zoll, Connell, & Grunebaum, 1986; Sharp et al., 1995). Thus, due to the higher levels of adversity typically present in LMIC settings, it was expected that the effect of postnatal depression would actually be larger, or at least more consistently found. Somewhat surprisingly, this was not the case. Rather, findings regarding the influence of postnatal depression on cognitive outcomes in LMICs appear to be characterised by the same inconsistencies as those from HICs. Although many of the studies controlled for variations in a variety of risk indicators within each sample, it was still expected that the overall level of risk relative to that found in a HIC would exacerbate any influence of postnatal depression. While it is not clear why this was not the case, it is plausible that the interaction between individual socio-economic risk factors and maternal depression functions differently in LMIC settings (Garman et al., 2019). It is also possible the distinct profile of risk in LMIC settings, particularly the severity and co-occurrence of risk factors (Britto et al., 2017; Herba et al., 2016), may be exerting such a strong independent effect on infant cognition that the unique influence of postnatal depression is not sufficient to predict differences in cognitive development.

It should also be noted that HIC studies have demonstrated that chronic depressive symptoms throughout infancy can have a greater impact on a number of domains infant development, including cognition, than a single postnatal episode in a variety of high- and low-risk cohorts (Brennan et al., 2000; Gueron-Sela et al., 2018; NICHD Early Child Care Research Network, 1999; Petterson & Burke-Albers, 2001). In the current review, only two studies examined chronic exposure in the postnatal period. While Garman et al. (2019) did not find an effect of chronic depression on development, Quevedo et al. (2012) found that only exposure to chronic depression predicted poorer language outcomes. This paucity and inconsistency highlights chronicity of exposure as an important area of research for future studies in LMIC settings.

2.5.1.3 Moderated effects

Although the main purpose of this review was to synthesise evidence for a main effect of perinatal depression there are a few studies which investigated moderated effects and are

worth mentioning briefly. Bandoli et al. (2016) found no main effect of antenatal depression but did find that female infants exposed to both peri-conceptional alcohol use and antenatal depression performed significantly worse than males exposed to the same. Black et al. (2007) also did not find a direct association between postnatal depression and change in cognitive skills from 6-12 months. However, there was a significant effect of the interaction between depression and infant temperament, with irritable infants of depressed mothers showing significant cognitive impairments, while easy-going infants did not. These results suggest that children with irritable temperaments may be at additional risk for poor development when their mothers are depressed. They also found that caregiving was significantly and negatively associated with maternal depression, and that it mediated the effect of the depression-temperament interaction on cognitive outcomes. This is consistent with HIC literature which has highlighted the dynamic interaction between maternal mental health, caregiving and infant temperament (Crockenberg & Leerkes, 2003; Murray et al., 2015). Finally, Ali et al. (2013) found a significant effect of the interaction between father's income and depression, with infants of depressed mothers whose fathers earned less than 3500 rupees per month significantly more likely to show language delay. This suggests that differing levels of socioeconomic risk within LMIC settings play a significant role in determining how detrimental perinatal depression is to infant development. Alternatively, paternal income may be an index of paternal education, with lower incomes reflecting lower levels of education. Parental education may be linked to language development through its impact on the level of cognitive stimulation in the home or more directly via heritability effects (Doyle et al., 2009; Kovas et al., 2005).

2.5.2 Methodological considerations and limitations of reviewed studies

2.5.2.1 Perinatal depression assessment

In this review, all but three studies used screening tools to assess maternal depression and this may have contributed to a lack of consistency in significant findings. This is because symptom questionnaires are more likely to also detect mild or moderate forms of depression, which may lead to an overall underestimation of effect of depression on development (Kingston et al., 2015). Clinical diagnostic tools focus on symptoms at a level that meet diagnostic threshold and are therefore more likely to represent significant impairment (Osborne et al., 2019). Unfortunately, there were an insufficient number of studies in this review that used a diagnostic interview to draw any conclusions as to whether infants of mothers with a clinical diagnosis of depression were any worse off than infants of mothers who were assessed using a screening instrument.

2.5.2.2 Cognitive development assessment

Most studies included in the review used measures which generated a standardised composite of cognitive functioning as the outcome. However, even between different versions of the same measure there may be stark differences in the contributing items. For example, the BSID-II composite includes items relating to language, while BSID-III has separate cognitive and language subscales. Three antenatal studies and one postnatal study examined both cognitive and language development separately using distinct subscales, while one study in each period looked exclusively at language development. Two studies employed specific measures of executive function rather than broader indices such as composite scores.

One pattern to emerge was the difference in results between studies using the 2nd and 3rd editions of the BSID. Across the whole perinatal period the 5 studies which used BSID-III found a significant effect of maternal depression on cognitive outcomes, while all 5 of the studies which used the older BSID-II did not. Interestingly, one of the primary justifications detailed for the changes made in BSID-III was the removal of certain items that may be biased in favour of certain racial or ethnic groups (Albers & Grieve, 2007). It is possible that these adaptations enhanced the cross-cultural validity of the instrument, leading to a greater sensitivity to changes in development in LMIC settings. This idea is further supported by the observation that the two studies that used cognitive assessments which had been developed specifically for the culture in which they were being used both report significant results (Ali et al., 2013; Patel et al., 2003).

2.5.2.3 Timing of exposure

Six out of the eight antenatal studies assessed exposure during the 3rd trimester, one did not clarify the specific time of assessment, and the other assessed exposure during the 2nd trimester. While the fact that five of the 3rd trimester studies found a significant effect does suggest that the 3rd trimester is a sensitive period for development, the lack of assessments at other points of pregnancy precludes conclusions about timing effects. More work is required in this area to investigate whether the effect of antenatal depression varies according when it occurs during foetal development.

There was no distinct pattern of results with regards to the timing of exposure in postnatal studies. Exposure was assessed at various points across the first 12 months and there is too much variability to conclude that any particular period within this timeframe significantly influences the effects of postnatal depression.

2.5.2.4 Confounding variables

There were no clear patterns that emerged in the data with regards to the key adjustment variables, with the exception that studies that controlled for more variables tended to report smaller effects. Interestingly, although not one of the key variables identified, only one study controlled for the effect of nutrition in their analyses. Nutrition is thought to play an important role in cognitive development and so studies in the future should seek to include it where possible. One of the major difficulties in comparing the findings of the different studies in this review is the wide array of adjustment variables included by different authors. The studies reviewed here varied significantly both in the number and nature of the variables that were controlled for in analyses. Even where the same variables were included, they varied in how they were measured or conceptualised. It is possible that these variations contributed to the inconsistency observed in the overall findings. Thus, a key recommendation from this synthesis is that it is important to control for a wide array of potentially confounding variables, but that researchers should strive to ensure consistency between studies in terms of the confounders included and how they are measured.

2.5.3 Review limitations

The main limitation of this review is the relatively small number of studies that were eligible for inclusion. The main reason for this is that perinatal mental health is very much an emerging field in many LMICs. As a result, one of the recurring themes of this review has been the statement that more studies are required in certain areas in order to draw firm conclusions. Bearing this in mind, the eligibility criteria for inclusion in the review were less narrow than may have been applied in a more developed field. One way in which the reach of the review was extended was to include RCTs under specific conditions. Although every effort was made to ensure that only findings that had not been confounded by the effects of any intervention were included in the final synthesis, RCT interventions may have had an undetected effect on cognitive development, thereby confounding results and potentially adding to the inconsistent findings. When only considering prospective cohort studies, all five antenatal studies and three out of six postnatal studies found a significant, negative main effect on cognitive outcomes. Although postnatal findings remain inconsistent when this restriction is applied, the case for an antenatal effect becomes contextually stronger.

Another potentially limiting factor is the broad categorisation of LMIC countries. While there are clear economic similarities within the countries designated as low- and middle-income included in this study, there are also very distinct cultural differences between them, which could influence the effect of perinatal depression on cognitive development.

However, the limited number of studies mean that it would not have been feasible to produce a systematic review for each country. Instead, while the heterogeneity amongst studies is not ideal, this review was able to draw together a previously disparate collection of studies and synthesise what is known and what is lacking in the existing literature. Similarly, cognitive outcomes are considered across a relatively heterogenous age ranges throughout infancy. Although it is possible that this may have affected results, there were no clear patterns in the findings to indicate markedly different effects at distinct ages. Additionally, the developmental period that spans infancy is often considered as uniform period of accelerated development (Black et al., 2017).

Although the risk of reporting bias is an ever-present in systematic reviews, the search strategy for this review included a systematic search of ProQuest for grey literature in the form of unpublished conference abstracts and dissertations. Additionally, the Newcastle-Ottowa Scale for Cohort Studies was used to assess study quality I number of domains, including participant selection and recruitment, sample attrition, and the methods and measures of exposure and outcome assessment. Key features identified in each study are given in Tables 1 and 2, and all but three studies received a quality rating of fair or good. A further limitation is that not all studies published mean scores for the depressed and non-depressed groups in their sample, and while some authors responded to requests for this information, others did not. This meant that even the most rudimentary comparison of unadjusted effect sizes could not be carried out. In the future, authors should ensure they present this basic descriptive information to aid comparison with other studies.

2.5.4 Implications for clinical practice and future research

The results of this review suggest that maternal depression during pregnancy can have a significant impact on infant cognitive development in LMIC settings. The importance of these findings are underlined by strong evidence from HICs for the continuity of cognitive ability from infancy to later life (Fagan, Holland, & Wheeer, 2007; Girault et al., 2018; Rose, Feldman, & Jankowski, 2011). The consequences of early cognitive impairments in LMIC settings might be even more wide reaching, perpetuating a cyclical pattern of lost potential and generational poverty (Daelmans et al., 2017). Despite an increase in the number of calls for the prioritisation of perinatal mental health in LMIC settings (Ganjekar, Thekkethayyil & Chandra, 2020), it is an area that has typically been neglected in these contexts (Baron et al., 2016). This review adds to the call for improved perinatal mental health services that will provide mutual benefits to the mother and the child. Research studies in this review demonstrated low adherence to the requirements for testing the independent effects of

maternal depression during the antenatal and postnatal periods. Given evidence from HICs that antenatal and postnatal depression exert independent effects on cognitive development, it is important that future studies are designed in a way that allows researchers to isolate the effects of each. This can be achieved in two ways. Firstly, studies which have explored trajectories of maternal depression or split the sample into groups based on episode timing across the perinatal period should be replicated with the purpose of isolating independent and cumulative effects of perinatal depression (Garman et al., 2019, Quevedo et al., 2012; Rotheram-Fuller et al., 2018). Alternatively, a study can control for the effects of maternal depression at other relevant timepoints, as demonstrated by Tran et al. (2013).

Given that the BSID-III was designed to minimise previous cultural bias, the observation that significant impairments were more consistently found when using the BSID-III as compared to the BSID-II suggest that effects may be sensitive to the choice of instrument. This thought is reinforced by the observation that the two postnatal studies which used outcome measures designed specifically for their population also found significant effects of depression (Ali et al., 2013; Patel et al., 2003) Therefore, an important step in developing the LMIC literature will be the selection or development of culturally sensitive measures of cognitive development. In addition, due to the existing reliance on screening tools, where possible future studies should incorporate designs which facilitate comparisons of the effects of the mild-moderate and more severe, clinical levels of depressive symptoms, namely.

Furthermore, while this review highlights the need for more high-quality studies investigating main effects, future studies should also look to explore a more complex relationship between perinatal depression and cognitive outcomes. In the light of recent advances in developmental science, investigations of a simple relationship between the two factors, even when controlling for a number of other variables, are inadequate (Sameroff, 2010). While it is clear that the level of risk is generally higher in LMIC settings, it is not clear how the relationship between risk and perinatal depression functions (Garman et al., 2019). Rather than simply controlling for possible confounding variables, studies need to investigate how variables such as infant temperament, parenting quality, infant gender and other key socio-economic indicators influence the relationship between perinatal depression and cognitive development (Ahun & Cote, 2018; Goodman & Gotlib, 1999).

More specifically, future studies need to be designed to answer questions arising from current theories and hypotheses in developmental psychology. A number of theories, such as the foetal programming hypothesis (Barker, 2004), the predictive adaptive response hypothesis (Bateson, Gluckman, & Hanson, 2014) and the differential susceptibility hypothesis (Belsky & Pluess, (2009) are increasingly influencing research in HICs and it is important for research LMICs to be similarly driven by current theory. The need for this is particularly evident in investigations of the role of postnatal depression, where findings for a main effect are very inconsistent. There is now considerable evidence that depression can compromise a caregiver's ability to respond sensitively and contingently to a child, and that a child's temperament can, in turn, influence the quality of the caregiver-child relationship and affect how susceptible they are to both positive and negative caregiving practices (Crockenberg Leerkes, 2003; Murray et al. 2015). Given the interactional, dynamic nature of the factors described above, it is apparent that testing only a simple, direct relationship could result in a significant underestimation of the effects of postnatal depression on child outcomes.

2.5.6 Conclusion

While based on a limited number of studies (n = 16), this systematic review gives valuable insight into what is known about the relationship between perinatal depression and infant cognitive development in LMIC settings. A key strength of this study is the focus on, and in-depth exploration of, the effect on cognitive development of perinatal depression alone. This contrasts with previous systematic reviews which have focused more broadly on a range of mental health disorders and neurodevelopmental outcomes, thereby retaining a high degree of novelty and adding considerable value to the existing literature. The findings provide a platform for the progression of research in this area by highlighting methodological and theoretical improvements that can be made to increase its relevance and impact.

Evidence for a main effect of antenatal depression on cognitive development was quite consistent, especially when restricted to prospective studies. There was some evidence that this effect is independent of postnatal depression and that effects endure into the child's second and third year. This provides a strong case for the inclusion of screening for maternal depression as part of routine antenatal care and for early intervention to be implemented in LMIC settings. Evidence for the direct influence of postnatal depression is more equivocal at present.

Importantly, findings from HICs suggest that the inconsistent findings regarding postnatal depression may be the result of its effect being evident in interaction with other factors. While there is some evidence of possible moderated effects of postnatal depression on cognitive development in LMICs, the majority of studies reviewed did not investigate these underlying pathways. The paucity of studies exploring anything more than a direct relationship between both postnatal and antenatal depression and cognitive development precludes drawing firm conclusions about this. More high-quality studies are needed in LMIC settings, driven by current theory, that test main effects and examine moderating or mediating pathways to child cognitive development.

Chapter 3: The prospective association between early and chronic postnatal depression and infant cognitive outcomes in India.

3.1 Background

There is now wide recognition that the time from conception until the child reaches 2 years of age is extremely important for development and particularly sensitive to external influences (Black et al., 2017). This may confer a developmental advantage or vulnerability depending on the environment and experience an infant is exposed to (Tiecher & Samson, 2016). A common adversity experienced during this period is postnatal depression, with a recent meta-analysis reporting a prevalence of 19% in India (Upadhyay et al., 2017). As well as predisposing mothers to chronic and recurrent episodes, postnatal depression represents a potentially significant adversity for the infant during a particularly vulnerable phase that may result in enduring developmental impairments (Fagan et al., 2007; Hoffman et al., 2017; Rose et al., 2011, 2012). The association between maternal postnatal depression and infant cognitive development has been extensively explored in high-income countries (HICs) but research in low- and middle-income countries (LMICs) is far more scarce, with only one study on the topic being published in India (Patel et al., 2003). The literature has been extensively reviewed in Chapters 1 and 2 so only key findings and studies are briefly summarised below.

As discussed in Chapter 1, in order to remain consistent and facilitate comparison with the existing literature, this project utilises an inclusive and focused conceptualisation of global infant cognitive development. Additionally, due to the hypothesised interdependent and inter-related nature of early cognitive and language milestones (Deak, 2014), both domains will be considered, reviewed and examined under the label of 'cognitive development'.

3.1.1 Early postnatal versus chronic exposure in HICs.

HIC research tends to conceptualise exposure to maternal depression during infancy in two ways, either as exposure in the period from immediate postpartum to the end of the infant's first year, referred to as an early postnatal, or "timing", effect or as continued exposure across an extended period of time during infancy, referred to as a "chronic" effect. In its most simple form, the research question being asked is whether the infant is more sensitive to the effects of maternal depression in the early postnatal period to the point where a single exposure during this time is equally or more detrimental to a child's development than exposure over a more extended period of time. It should be noted that performing a stringent test of a "sensitive period" is extremely challenging and many of the

studies reviewed did not do this. Rather, they offer useful indications as to whether such a putative sensitive period for the impact of maternal depression on development may exist.

While there is some evidence for a timing effect of maternal depression in HICs, findings are inconsistent. A number of studies did find a significant effect of depression during the first year postnatal (Bornstein et al., 2012; Conroy et al., 2012; Kawai et al., 2017; Koutra et al., 2013; Lyons-Ruth et al., 1986; Murray, 1992; Perra et al., 2015; Smith-Nielsen et al, 2016; Whiffin & Gotlib, 1989), but a similar amount did not (Hanley, Brain & Oberlander, 2013; Kaplan et al., 2014; Keim et al., 2011; Kiernen & Huerta, 2008; Piteo, Yelland & Makrides, 2012; Stanley, Murray & Stein, 2004; Stein et al., 2008). There is, however, much stronger evidence to suggest that chronic exposure to depression during infancy can impair cognitive development, with the majority of studies investigating chronic exposure finding a significant effect (Azak, 2012; Brennan et al., 2000; Cornish et al., 2005; Milgrom et al., 2004; Sutter-Dalley et al., 2012; NICHD Early Care Research Network, 1999; Park et al., 2018). Only one study found evidence suggesting that the timing of the onset of depression was more important than the length of exposure to depression. Sharp et al. (1995) reported that, for male infants only, exposure during years 1-4 added no further impairment when the child had been exposed to depression in the first year and was not associated with impairment when there was no depression reported in the first year.

3.1.2 Maternal sensitivity as key mediator/moderator.

A growing body of research has found evidence that sensitive caregiving can be compromised by maternal depression (Bernard et al., 2018) and a number of large-scale studies have found evidence that impaired sensitivity to non-distress may be a key pathway through which maternal depression exerts its effect on child cognitive development, either as a mediator or a moderator (Gueron-Sela et al., 2018; Milgrom et al., 2004; NICHD Early Care Research Network, 1999). There is also evidence suggesting that male infants are more at risk from maternal depression and impaired maternal sensitivity (Azak, 2001; Hay et al., 2001; Murray et al., 1993, 1996, 2010; Sharp et al., 1995), alongside evidence regarding the roles of a number of additional factors involved in influencing development, such as antenatal depression (Glover, 2014), socioeconomic status (Kurstjens & Wolke, 2001), maternal IQ and education (Walker et al., 2011), maternal age (Barnes et al., 2014) and parity (Bjerkedal, 2007).

3.2.3 Existing findings from LMIC literature.

The systematic review of studies conducted as part of this thesis returned results from 9 studies conducted in LMICs. The review was restricted to studies which explored the effect of early postnatal depression. Similar to the HIC literature, findings regarding a timing effect were inconsistent, with four studies finding a significant effect of maternal depression in the first year postnatal (Ali et al., 2013; Galler et al., 2000; Patel et al., 2003; Quevedo et al., 2012) and five studies finding no effect (Black et al., 2007; Familiar et al., 2018; Garman, Cols, Tomlinson, Rotheram-Borus, & Lund, 2019; Hamadani et al., 2012; Tran et al., 2013). The study by Patel et al. (2003), which did find a significant effect of postnatal depression on cognitive development at 6 months, was the only one of these studies conducted in India. More recently, Thomas et al. (2020) reported finding no significant effect of antenatal depression on cognitive or language outcomes at 30 months after adjusting for covariates. Although focusing on a different part of the perinatal period, the similarity of the sample, also drawn from urban Bangalore, provides important context for the current study.

Although detailed discussion of any effects beyond the first year postnatal was outside the scope of the systematic review, it is worth noting that Garman et al. (2019) did investigate the effects of chronic depression beyond the first postnatal year on cognitive outcomes. Depressive symptoms were modelled as four trajectories across three timepoints, including 2 weeks, 6 months, and 18 months. Trajectories were chronic low, early postpartum (symptoms peaked at 6 months and then decreased), late postpartum (symptoms low at 6 month and sharp increase at 18 months) and chronic high (symptoms high at each time-point). However, the authors did not find a significant effect of any of the trajectories on cognitive outcomes assessed using the Bayley Scales of Infant Development at 18 months or an executive function battery at 36 months Other than that, there are no studies that the author is aware of that investigate the effects of chronic exposure to maternal depression on infant cognitive development conducted in LMICs. The majority of studies synthesised focused on the main effect of depression and only a few explored more complex moderated effects. Only one study considered the role of caregiving (Black et al., 20007) and two investigated gender effects (Bandoli et al., 2016; Donald et al., 2019). Furthermore, the majority of studies did not control for the effects of antenatal depression, which is a significant predictor of postnatal depression (Rahman & Creed, 2007) and has been found to have an independent effect on cognitive development operating through a distinct mechanism (Osborne et al., 2018).

In addition to the covariates listed above that have been highlighted by the HIC literature, shared caregiving may play an important role in infant development in the Indian context. While the responsibility of caregiving falls primarily on the mother in western contexts (Goodman & Gotlib, 1999), in India, which is considered to be a collectivist culture and where extended or joint families are far more prevalent, other caregivers such as fathers, grandparents, other family members, and even neighbours, are likely to play a more important role in child-rearing (Breton, 2019; Chadda & Deb, 2013). Although multiple caregivers are often involved in child-rearing in western contexts, their role tends to be lesser and is mainly ordered in a serial manner, with one caregiver "swapping" in for another at defined points. In contrast, shared caregiving in non-western settings is much more fluid and multiple caregivers are often present at the same time, taking turns to respond to the infant with no clear place- or time-bound task division (Mesman, Minter & Angnged, 2016). This increased frequency and distinct pattern of shared caregiving may offer a unique contribution to infant development, either positive or negative, and so it is important to consider. Although shared caregiving is not the main focus of the current study it is likely to play an important role and so a basic measure pertaining to the presence or absence of an alternate caregiver will be included as a covariate. This provides an important piece of contextual information that will be considered alongside other factors that may be operating in a unique manner in this context, such as an increased level of adversity present in LMICs and potential issues of gender bias in India in favour of a preference for a male infant (Barcellos, Carvalho, & Lleras-Muney, 2014; Britto et al., 2017).

3.2.4 Aims of the current study.

The aim of this study is to further advance the literature regarding the relationship between maternal depression and cognitive development during infancy, within the general LMIC context and, more specifically, within India. As indicated in Chapter 1, the current thesis uses a broad definition of early cognitive development that also includes language development. Thus, the term "cognitive development" will be used throughout to refer to the overarching focus of this study, but results will also be discussed more specifically in terms of the cognitive and language subscales utilised by Bayley Scales of Infant Development. The study aimed to examine (i) the effect of exposure to depression during the early postnatal period (8 weeks) and (ii) the effect of chronic exposure over the first two years of life, while controlling for a number of relevant covariates, including antenatal depression. A further aim of the study was to explore sex-specific effects and investigate the role of maternal sensitivity to non-distress in this relationship.

3.2.5 Study hypotheses.

1. Exposure to early maternal postnatal depression (8 weeks) will be associated with impairments in infant cognitive development at age 2.
2. Chronic exposure to maternal depression will be associated with significant impairments in cognitive development at age 2. This association will be stronger than that of early postnatal depression.
3. Postnatal and chronic exposure to maternal depression will be associated with impaired cognitive development via, or in interaction with, maternal sensitivity to non-distress, and this will be more pronounced in male infants.

3.2 Method

3.2.1 Sample.

3.2.1.1 Study design

Participants were mothers and infants taking part in the Bangalore Child Health and Development Study (BCHADS), a prospective longitudinal cohort study starting in pregnancy and designed to study early risk and protective factors for childhood mental health throughout pregnancy and infancy, in India. BCHADS was established as a continuation of the Prospective Assessment of Maternal Mental Health Study (PRAMMS). Although now considered to be one unified study and typically referred to under the name BCHADS, the two studies will be described separately here to give clarity to the recruitment process.

The PRAMM study began in October 2014 and included planned follow-up assessments up to 8 weeks postpartum with the aim of examining prenatal mental health and psychosocial risk for preterm delivery and low birth weight in India. In 2016, as the result of a collaboration between NIMHANS and the University of Liverpool, the PRAMM study was extended to include a wider array of assessments up to age 2 and became BCHADS. The study was designed to mirror the Wirral Child Health and Development Study (WCHADS) wherever possible to allow for parallel measurement, but also included a variety of additional measures to ensure comprehensive coverage of the risk and protective factors unique to the Indian context.

3.2.1.2 BCHADS Ethics Statement

Ethical approval for phases 1 to 5 of BCHADS was granted by the National Institute of Mental Health and Neurosciences (NIMHANS) Institutional Ethics Committee (IEC) on 22nd June 2013. These phases were solely funded by the ICMR and so did not require UK ethics

approval. Ethical approval for phases 6 to 9 was granted by the NIMHANS IEC on 2nd July 2015 and by the University of Liverpool ethics committee on 1st March 2016. The letters confirming ethical approval for these phases of study are in Appendix 1, Participants gave written informed consent for data collection at multiple phases within the BCHADS. Information sheets that are relevant to the current thesis are given in Appendix 2.

3.2.1.3 Study Setting

Bangalore had an estimated population of 12.34 million in its urban area in 2017 during the study period, up from 8.5 million in 2011. It is now the 24th most populous city in the world and has a population density of four thousand per square kilometre. The main languages spoken include Kannada (38%), Tamil (28%), Telugu (17%), Urdu (13%), Malayalam (3%), Hindi (2.5%). About 79% of the population is Hindu, in line with the Indian national average. Muslims account for 13% of the population, followed by Christians (6%) and Jains (1%). About 10% of residents live in slums. Much of the population growth in Bangalore is due to migration from other states. Bangalore Urban also has a very skewed female-male gender ratio: 907 women for every 1,000 men, and has the lowest work participation rate among women, with just 24% of women working. 50% of Karnataka residents, including those living in rural areas, possessed a government Below Poverty Line card for providing benefits like social security and health care (National Family Health Survey, 2015).

The city of Bangalore is divided into nine zones, which are further subdivided into a total of 198 wards. Permission was obtained from the local municipality, Bruhat Bangalore Mahanagara Palike (BBMP), for the recruitment of participants who were registered at the Banashakari Urban Primary Health Centre (UPHC), which serves as the Government Referral Hospital in the south zone of the Bangalore Urban District. This site was chosen because it caters to women from low resource settings and is also reachable within an hour from the main study site (NIMHANS).

3.2.1.4 Recruitment Strategy

Recruitment to the study took place between July 2014 and May 2016 for the PRAMM study, and between May and November 2016 for the additional mothers included in the BCHAD study. Expectant mothers living in low-income areas of urban Bangalore were eligible for inclusion in the study if they were registered at the Banashankari UPHC. A consecutive sample of expectant mothers were approached by research assistants when attending appointments at three community antenatal clinics during their first or second trimester and screened for inclusion in the study. Changes in the healthcare system meant

that the antenatal visit in the first trimester became non-mandatory from May 2016, and so after this date many women could be approached only in their second trimester.

Screening was conducted by the team of research staff and women who met the criteria were invited to participate in the study. Women who had a major mental illness such as psychosis or a bipolar disorder, who were identified to have major health complications during the current pregnancy or were currently using alcohol or other psychoactive substances were excluded from the cohort.

Informed consent was sought and women who consented to be part of the study were recruited to the cohort. Out of the 1048 women approached, 909 women consented to participate, 698 at trimester one and 211 trimester two. 84 of these mothers were excluded at later time-points in the study: 44 because the pregnancy was high risk or had a poor outcome (e.g., miscarriage, medical termination, intrauterine death), 5 because of still birth, 24 because of neonatal death (under one month), 6 because of infant death (under one year), and 5 because infants were twins. Therefore, 825 women with a mean age of 23.02 years (SD = 3.5, range 17-39) remained eligible for postnatal follow-up. Full details of the sampling process and rate of attrition up to 2 years postpartum are illustrated in Figure 3.2.1.4.

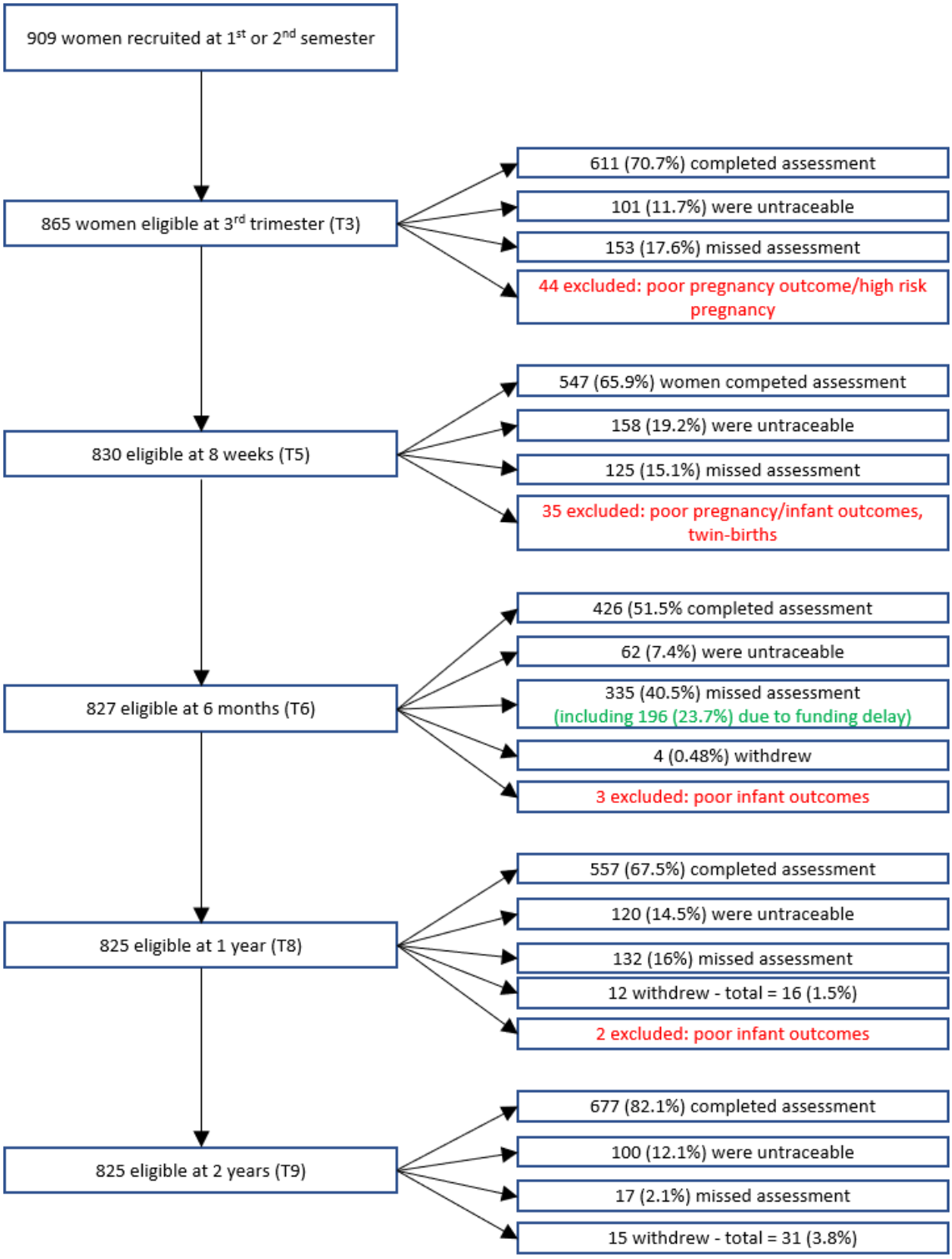
3.2.1.5 Assessment phases.

Assessments reported as part of this project represent a subgroup of a wider assessment battery used in BCHADS. Participants were assessed in the antenatal period during trimester one (T1: maternal questionnaire), trimester 2 (T2: maternal questionnaire), and trimester 3 (T3: maternal questionnaire). Antenatal assessments were conducted at the Banashankari UPHC, at the Anganwadi centres in the women's localities, or in their home, based on the participants' preference.

After delivery (T4) women received a telephone call from the research assistants to collect data about the pregnancy outcomes. Some of the birth outcomes, such as live births, Medical Termination of Pregnancy (MTP), miscarriages, Intrauterine Deaths (IUD), neonatal deaths and still births, and birth weight were recorded based on medical records.

Postnatal assessments took place at 6-12 weeks postpartum (T5: maternal questionnaire), 6 months postpartum (T6: maternal questionnaire, mother-child observation), 12 months postpartum (T8: maternal questionnaire, mother-child observation), 24 months postpartum (T9: maternal questionnaire, infant cognitive development). Postnatal follow-ups took place at the NIMHANS Centre for Wellbeing (NCWB) or at home.

Figure 3.2.1.4: Flow chart illustrating the rate of participants and attrition at each follow-up.



3.2.1.6 Sociodemographic characteristics of the cohort at recruitment.

Of the 825 families who were eligible for follow-up from birth 742 families (89.9%) gave data at least once postpartum. The general characteristics of the 83 women who were initially recruited, but who were lost to follow up in the postpartum were similar to those of

the 742 women who gave data postnatally (Table 3.2.1.6). The two groups of mothers were similar on all key demographic details except religion and family type. A significantly higher proportion of included mothers reported being Hindu and belonging to joint families. Although not significant, the inadequate uptake of antenatal care in the excluded sample was substantially higher, indicating that these mothers were generally less engaged with healthcare services. Importantly, there was no difference in mean depression at baseline between the two groups. The mean age of the mothers who provided postnatal data was 23.04 (SD = 3.45), 69.7% were educated only up to secondary level, and 11.2% were classed as below the poverty line in terms of their socio-economic status.

Compared to the urban Karnataka population whose data are provided by the 2015-16 National Family Health Survey (NFHS-4) (IIPS and ICF, 2017), the present cohort is similar in terms of education (72% of Karnataka women age 15-49 are literate - i.e. have either completed at least standard six or passed a simple literacy test) and religion (75.2% woman in the urban areas and 91.3% in the rural areas are Hindu). They were also similar in the proportion who were living below the Poverty line (13% of women living in urban areas of Karnataka are below the Poverty line). However, women from the BCHADS cohort were less likely to be employed at recruitment in pregnancy compared to the proportion employed in the general female population of Karnataka (35% aged 15-49).

Table 3.2.1.6: Sociodemographic profile of the BCHADS cohort

Characteristic	**Postnatal assessment** (at least one between 8 weeks and 2 years) **(N= 742)**	**NO postnatal assessment** (data provided only in pregnancy) **(N= 83)**	***p***
MOTHER			
Age (M, SD)	23.04 (3.450)	22.92 (3.663)	ns
Education (Up to Secondary)	69.7%	75.9%	ns
SES (LSES/BPL)	11.2%	11.6%	ns
Occupation (Employed)	13.8%	20.7%	ns
Religion			
Hindu	84.1%	75.6%	
Muslim	15.2%	20.7%	.012
Christian	0.7%	3.7%	
EPDS Depression symptoms at recruitment (M, SD)	2.67 (5.047)	2.91 (4.657)	ns
Parity (Primiparous)	45.5%	48.2%	ns
Antenatal care (Inadequate)	34.0%	46.2%	ns
Family Type			
Nuclear	41.4%	56.6%	
Joint	49.7%	33.7%	.018
Extended	8.9%	9.6%	
CHILD			
Gender (Male)	50.1%	-	-
Term (37-42 weeks)	80.4%	-	-
Birth weight - Kg (M, SD)	2.90 (0.534)	-	-

Note: - = postnatal data not available; p values calculated using T-tests for continuous variables and $\chi 2$ for categorical variables.

3.2.1.7 Assessments used in the current study and missing data.

This report contains data from women who completed key questionnaires and assessments at each assessment phase. Participants were only included if they had full data on each relevant measure. The only exception to this was the maternal depression assessments. For the creation of the antenatal depression variable, in order to maximise data, depression was coded as present if it had been recorded at any time-point in the antenatal period. Additionally, maximum likelihood estimation was used to account for any missing time-

points in the postnatal maternal depression assessments in the creation of the chronic depression factor.

3.2.2 Measures

3.2.2.1 Maternal measures

Maternal Depression. Maternal depression was assessed at each phase using the Edinburgh Postnatal Depression Scale (EPDS) (Cox, Holden, & Sagovsky, 1987). The EPDS is a 10-item Likert scale self-report instrument designed to detect postnatal depression by focusing on the cognitive and affective aspects of depression rather than somatic symptoms. It has been widely used and validated in a number of settings (Eberhard-Gran et al., 2001; Murray & Carothers, 1990). Each item addresses a distinct symptom of postnatal depression and is rated using a unique set of response items, scored 0-3. The instrument gives a total score of 0-30, with a higher score indicating greater distress. A cut-off of ≥13 was validated for probable depression by Cox et al. (1987) and is routinely used in western settings. However, when applied in different cultural contexts wide variation in psychometric properties has been reported, indicating that valid threshold may vary considerably between contexts (Gibson, McKenzie-McHarg, Shakespeare, Price, & Gray, 2009). Tran et al. (2002) report that thresholds are generally lower in low income settings, and themselves found an optimal cut-off of ≥3 in the perinatal period in Vietnam.

For the purposes of the current study the EPDS was translated into the local language (Kannada) following World Health Organisation guidelines for measure translation. Forward translation was done by a researcher from the team who was a native speaker of Kannada with bilingual proficiency in Kannada and English. The translated version was then reviewed by experts in the field to ensure the appropriateness of the terms and phrases used in the questions and revisions were made based on their suggestion. This version was then back-translated to English by a translator with bilingual proficiency (see Appendix 4). The original English questionnaire, the translation in Kannada and the back-translated version were compared and reviewed to identify any differences and appropriate modifications were made. Finally, a pilot study was conducted with 10 pregnant mothers from the target population. Final modifications were made to the phrasing of items 1, 2, 4, 5 and 6 based on pilot feedback (see Appendix 4). The measure was then tested against diagnoses of major depressive disorder using the Mini International Neuropsychiatric Interview (Lecrubier et al., 1997). Data suggested the existence of distinct reporting styles between the antenatal and immediate postnatal period (8 weeks) and later in infancy (6-24 months),

so two separate cut-off scores were validated using independent samples. A cut-off of ≥3 (sensitivity = 62.5%, specificity = 73.9%) was selected for phases T1-T5 and a cut-off of ≥10 (sensitivity = 95.2%, specificity = 83.1%) was selected for phases T6-T9. The full details of the ROC analyses are reported in Appendix 4.

Early postnatal depression was measured using the T5 EPDS data only. Those scoring above the cut-off were identified as depressed and the data was converted into a binary variable (0 = non-depressed, 1 = depressed) for use in analysis. The chronic depression exposure was measured by creating a factor score from categorical EPDS data at T5, T6, T8 and T9, using the relevant cut-off at each timepoint.

Maternal Sensitivity. Maternal sensitivity, defined as the mother's ability to perceive child signals, interpret these signals correctly, and to respond to these signals promptly and appropriately (Ainsworth, Bell, & Stayton, 1974), was assessed using a standard 15-minute semi-structured play-based procedure adapted from the NICHD mother-infant interaction scales at either the 6-month or 12 -month assessment, and coded using the "Qualitative Ratings for Parent-Child Interaction at 3-15 Months of Age" (Cox & Crnic, 2003). During the first 7-minutes, mothers were asked to play with the infants using a familiar toy or object. For the remaining 8-minutes, mothers were given a standard set of toys to play with. Toys for the standard set were selected to be broadly parallel to the standard set used in western settings but familiar to mothers and infants in India. Assessments were recorded at the NCWB unless the participants did not want to attend the centre, in which case they were filmed at home. NCWB assessments were filmed using two static video cameras placed at different angles to capture facial expressions and interactions as the dyad moved around the room. A play mat was positioned in the optimal field of view for the cameras and mothers were asked to remain on it where possible. When assessments were conducted in at home, researchers recorded assessments using a single portable video camera and tripod.

The NICHD coding scheme provides a wide range of observer-rated dimensions of caregiving dimensions, and previous approaches have included creating a composite or factor score from multiple scales (NICHD Early Child Care Network, 1999, Gueron-Sela et al., 2018) or have examined the influence of single scales (McElwain & Booth-LaForce, 2006). In this study, only the sensitivity to non-distress rating was used as an indicator of maternal sensitivity as this is believed to be most relevant to cognitive development (Cecil et al., 2017; Murray et al., 2008). Sensitivity to non-distress (SND) captured the extent to which a

mother noticed and responded appropriately to the social gestures, expressions, and signals of the infant, and how well timed, paced and child centred the interaction was. Each scale is coded on a 5-point global impression scale from 1 (not at all) to 5 (highly characteristic). Videos were coded by two raters from the local research team who were trained to reliability in the coding scheme by a gold standard NICHD rater (IRRs = .86-.89). All raters were blind to the depression and socio-demographic status of the mothers. The coding scheme for SND can be seen in Appendix 5.

Covariates: Covariate data were obtained from a socio-demographic questionnaire administered at baseline and included socio-economic status (SES), maternal age, maternal education, parity and infant gender. In order to control for antenatal exposure to depression and to maximise the data available, a dichotomous variable was created that merged data from each antenatal timepoint and represents caseness at any point. SES was derived from family income and mothers were classified *Below Poverty Line/Low-SES (<6000 INR per month), Higher Low-SES (HLSES, 6000-11,999 INR per month), Middle-SES (MSES, 12,000-30,999 per month)* or *Higher-SES (HSES, ≥32,000 INR per month)*. This was then collapsed into a dichotomous variable representing *BPL/HLSES (55.9%) (1)* vs *MSES/HSES (44.1%) (0)* for analyses. Maternal education was classified as the highest qualification obtained from the following: *Uneducated (4.7%), Primary (13.5%), Secondary (52.8%), Pre-University Course (18.3%), Undergraduate (9.3%), Postgraduate (0.9%), Professional Course (0.4%)*. This variable was then collapsed into a dichotomous variable, *Up to Secondary (0)* vs *Above Secondary (1).* Parity was defined as whether the mother was *primiparous (0)* or *multiparous (1)* at the time of consent. Data regarding presence of an alternate caregiver (ACG) was obtained from a shared caregiving checklist, which requires the participant to indicate whether there are any secondary caregivers were involved in child-rearing. Data from T6 and T8 was combined to create a dichotomous variable that indicated whether there was an ACG *present (1)* or *absent (0)* at either postnatal timepoint.

3.2.2.2 Child measures

Child Cognitive Development: Child cognitive and language development was assessed at 24 months using the Bayley Scales of Infant Development (Third Edition) (BSID-III) (Bayley, 2006). Although it has not been validated in an Indian population, the BSID-III is considered as a gold-standard measure of infant development globally and has been widely used in low- and middle-income settings (Breen et al., 2018). The tool measures development by direct observation across a number of domains, including *cognition, receptive language* and *expressive language.* The cognitive scale assesses areas including sensorimotor

development, exploration and manipulation of objects, object relations, concept formation, memory and problem solving. The expressive language scale assesses preverbal communication, vocabulary use and morpho-syntactic development, and the receptive language scale assesses the child's ability to recognise sounds and the development of verbal comprehension and vocabulary. These scales were administered in the local language and scored by research assistants who had been trained by a certified trainer from Pearson Academy India. Further training was provided by a clinical psychologist with experience using the BISD-III. Research assistants were provided with a video recording of the full assessment administered by the clinical psychologist and then conducted two assessments under his supervision. Following this, five assessments administered by each research assistant were randomly selected to establish inter-rater reliability in relation to the domain specific test scores and overall scores. Only when there was 100% agreement on the domain specific score and less than 5% of uncertainty in interpreting individual items, the research assistants were permitted to carry out the assessments independently. Random checks were then carried out throughout the study to ensure assessments were being administered to the required standard.

Following training, each researcher conducted a video-recorded pilot assessment and was given feedback on administration and scoring. Raw scores for the cognitive, receptive language and expressive language subscales were converted into norm-based scale scores (range 1-19, *M* = 10, *SD* = 3), and then into two composite scores (*M* = 100, *SD* = 15), one for the scaled cognitive score and one for the sum of the two language scaled scores. As no Indian norm scores for the BSID-III exist, American norms were used for the scoring, a practice that has been commonly adopted in the LMIC literature (Breen et al., 2018; Quevedo et al., 2012; Tran et al., 2012)

3.2.3 Analysis plan

First, in order to create the chronic postnatal depression factor score, the tetrachoric correlations between the EPDS data at T5, T6, T8 and T9 were examined and the data were modelled as a single latent variable (Chronic Depression) using the gsem command in Stata version 14 (Statacorp, 2015). A factor score was then extracted and used for all subsequent analyses. This approach uses *maximum likelihood estimation* to create a factor score for each participant that represents the number of depressive episodes present across the four timepoints, with a higher score representing more episodes of depression. This approach is able to deal with missing data in a robust manner by using the available data to predict missing responses.

All subsequent analyses were conducted using Statistical Package for Social Sciences (SPSS) version 24 for Windows. Continuous variables were checked for normality and transformed where necessary. Following transformation, only chronic depression remained marginally skewed but was considered acceptable for use as a predictor in multiple linear regression.

Analyses were conducted independently for testing the associations between cognitive development and exposure to early postnatal and chronic depression, and conducted using restricted samples consisting of participants who had complete data for all the included variables. Bivariate associations were examined using Pearsons correlations and, where appropriate, Spearmans or Tetrachoric correlations. T-tests were also conducted to examine differences in outcomes by categorical predictors. Mediation analyses were planned in line with study hypotheses but could not be tested due to the lack of main effects. Moderation analyses were carried out as planned. Multiple linear regression models were constructed separately for each outcome, first in using data from the full sample, and then split by gender. Predictors were entered in a series of blocks, with each model adhering to the same basic structure. Maternal depression was entered first to test its individual unadjusted contribution to development. The second step added the effects of the covariates. The third step tested the main effect of sensitivity to non-distress. In the final step, the Maternal Depression X Sensitivity interaction term was added. All continuous variables were centred prior to creating the interaction terms.

3.3 Results

3.3.1 Descriptive statistics.

Mothers in this study reported rates of probable depression at each time-point that were broadly similar to those reported in various meta-analyses (Shorey et al., 2018; Upadhyay et al., 2017; Woody et al., 2017). At 8 weeks postpartum, using a cut-off of ≥3, there was a prevalence of 19.3%. At 6, 12 and 24 months, using a cut off of ≥10, there was a prevalence of 15.4%, 15,8% and 14.4%, respectively. Table 3.3.1 shows descriptive statistics for key study variables and present a comparison between the full cohort and the restricted samples for the early postnatal and chronic exposure analyses. The early postnatal sample was restricted to participants who provided EPDS data at 8 weeks postnatal (T5) and provided data for each of the other relevant variables (n = 309). The chronic sample was restricted to participants who provided EPDS data for at least 2 postnatal assessments and provided data for each of the other relevant variables (n = 395). The chronic sample was made up of all 309 participants that made up the early postnatal sample and 86 additional

participants who met the criteria for the chronic depression factor of providing EPDS data for at least other timepoints. The restricted samples were broadly similar to the full cohort that provided data for at least one postnatal assessment. However, in both the early postnatal and chronic exposure subsamples, mothers were significantly older, and a higher proportion reported experiencing depression in the antenatal period. A higher proportion of mothers in both subsamples also reported that an ACG was present at 6 or 12 months.

3.3.2 Bivariate analysis

Bivariate associations are presented below. Correlation coefficients will be interpreted as follows: <.10 = very weak; .10-.39 = weak; .40-.69 = moderate; .70-.89 = strong; .90-1.00 = very strong (Schober, Boer, & Schwarte, 2018).

3.3.2.1 Early postnatal sample

Bivariate associations are presented in Table 3.3.2.1a for the early postnatal sample, and in Table 3.3.2.1b for boys and girls separately. Note that correlations between categorical variables are tetrachoric, which are typically larger than Pearson correlations and thus may not be significant at the same magnitude. In the whole early postnatal sample, there was no association between early postnatal depression and cognitive development but there was a significant but weak negative association with language development. Postnatal depression was also positively associated with antenatal depression and parity, indicating that mothers with higher EPDS scores at 8 weeks were more likely to report depression during pregnancy and more likely to be multi-parous, and negatively associated with maternal sensitivity and maternal education, indicating that mothers with postnatal depression had lower sensitivity scores and were less likely to have completed education beyond secondary level. Cognitive development was not associated with any of the covariates, but language development was negatively associated with parity, indicating lower scores for infants who were not first-born, and positively associated with gender, indicating that girls typically had scores higher than boys. There was also a positive association between maternal age and parity, indicating older mothers were more likely to be multi-parous, and a negative association between education and child gender, indicating that mothers of females children were less likely to have completed education beyond secondary level. Finally, there were marginal negative associations between the presence of an ACG and both antenatal and postnatal depression, indicating that it was less likely for mothers to be above the cut-off for depression when an additional caregiver was present.

Table 3.3.1: Comparison of restricted samples used for analyses against full cohort on key study variables

Measure	Full Cohort	Early Postnatal Sample (n = 309)	*p*	Chronic Sample (n = 395)	*p*
Maternal Age (M, SD)	23.04 (3.45)[a]	23.44 (3.65)	.001	23.42 (3.63)	.007
Education (Up to secondary)	69.7%[a]	69.9%	ns	68.1%	ns
SES (BPL/Upper Lower)	56.5%[b]	57.3%	ns	55.9%	ns
Parity (Primiparous)	45.5%[a]	42.4%	ns	44.3%	ns
ACG (T6 or T8) (Present)	76.0%[c]	79.9%	.008	79.2%	.018
Antenatal Depression (Depressed)	37.5%[d]	42.7%	.036	41.0%	.013
T5 Depression (Depressed)	19.1%[e]	21.4%	ns	21.4%	ns
Chronic Depression (M, SD)	-0.003 (0.88)[d]	0.03 (0.99)	ns	0.002 (0.96)	ns
Sensitivity to Non-Distress (M, SD)	2.71 (0.99)[f]	2.71 (0.96)	ns	2.72 (0.99)	ns
Child Gender (Male)	49.4%[g]	53.4%	ns	51.9%	ns
BSID Cognitive (M, SD)	93.90 (8.29)[h]	93.77 (8.29)	ns	93.71 (8.24)	ns
BSID Language (M, SD)	98.34 (11.72)[i]	97.57 (11.67)	ns	97.62 (11.59)	ns

Note – [a] n = 739; [b] n = 616; [c] n = 587; [d] n = 741; [e] n = 538; [f] n = 513; [g] = 733; [h] n = 661; [i] n = 660; p values calculated using t-tests for continuous variables and $\chi 2$ for categorical variables.

When bivariate associations were calculated for boys and girls separately, neither group showed any significant associations between early postnatal depression and either cognitive or language development. For boys, postnatal depression was negatively associated with maternal education, as it was in the whole sample, and positively associated with SES. As SES is reverse coded, this indicates that mothers with a lower SES were more likely to report higher levels of postnatal depression. The marginal negative association between postnatal depression and the presence of an ACG became significant for boys, indicating that lower depressive symptoms were associated with the presence of an ACG. Language was positively associated with maternal sensitivity, suggesting that more sensitive early caregiving was associated with better language outcomes in boys. Cognitive development was not associated with maternal sensitivity but was negatively associated with SES (reversed) and the presence of an ACG, and positively associated with maternal age. This suggests boys with older mothers or who were in a higher SES performed better, while they tended to perform worse in the presence of an ACG. Additionally, a positive association was observed between maternal sensitivity and education, and a negative association was observed between antenatal depression and maternal age. For girls, both cognitive and language development were negatively associated with parity, indicating that higher scores on both were associated with being the first-born child. As in the whole sample analysis, positive associations were observed in boys and girls between cognitive and language development, maternal age and parity, and between postnatal and antenatal depression.

Table 3.3.2.1a: Bivariate associations between key study variables in early postnatal sample

	Language	Early PND[a]	AND[b]	SND[c]	Maternal Age	Education	Parity	SES	ACG[d]	Gender
Cognitive	**.51****	-.05	-.07	.04	.07	-.03	-.06	-.10	-.01	.01
Language		**-.12***	-.09	.06	.03	.00	**-.11***	-.04	.00	**.11***
Early Postnatal			**.50*****	**-.13***	-.05	**-.30****	**.20***	.12	-.21†	-.08
AND				.02	-.09	.01	-.13	.09	-.19†	-.03
SND					.03	.04	-.02	.00	.04	.05
Maternal Age						.00	**.46****	-.11	-.01	.06
Education							-.18†	-.17†	.02	**-.20***
Parity								.04	-.13	.06
SES									-.10	.07
ACG										-.15

Note – [a] Postnatal Depression; [b] Antenatal Depression; [c] Sensitivity to Non-Distress; [d] Alternate Caregiver; * p < .05; ** p < .001; Tetrachoric correlations reported for associations between categorical predictors.

Table 3.3.2.1b: Bivariate associations between key variables in early postnatal sample separately for boys and girls. Boys (blue) in the top diagonal, girls (green) in the bottom diagonal.

	Cognitive	Language	Early PND[a]	AND[b]	SND[c]	Maternal Age	Education	Parity	SES	ACG[d]
Cognitive		**.51****	-.04	-.02	.08	**.16***	-.05	.06	**-.20****	**-.17***
Language	**.53****		-.14	-.11	**.18***	.09	-.01	.00	-.15	.00
Early PND	-.07	-.09		**.41****	-.15	-.04	**-.36***	.14†	**.26***	**-.32***
AND	-.13	-.07	**.61*****		.01	**-.20***	-.01	-.24†	.13	-.23
SND	.00	-.05	-.10	.03		.09	**.18****	-.01	-.03	.07
Mat Age	-.05	-.05	-.06	.02	-.03		.07	**.45****	-.10	-.04
Education	.00	.04	-.24	.02	-.08	-.08		-.18	-.20	.06
Parity	**-.20***	**-.25****	.29†	.01	-.03	**.41****	-.16		.05	-.21
SES	.04	.05	-.05	.05	.03	-.13	-.11	.03		-.22
ACG	**.17***	.02	-.11	-.16	.02	.03	-.05	-.04	.02	

Note – [a] Postnatal Depression; [b] Antenatal Depression; [c] Sensitivity to Non-Distress; [d] Alternate Caregiver; * p < .05; ** p < .001; Tetrachoric correlations reported for associations between categorical predictors.

3.3.2.2 Chronic sample

Bivariate associations are presented in Table 3.3.2.2a for the chronic sample, and in Table 3.3.2.2b for boys and girls separately. As many of the associations were similar between samples, only differences between associations for the chronic and early postnatal datasets will be highlighted. Generally, chronic depression showed fewer associations with other variables. Unlike the early postnatal sample, there was no association between chronic depression and language development. There was also no association between chronic depression and maternal sensitivity, maternal education or parity. Associations between language development and other variables were generally similar, but there was an additional association between, cognitive development and SES, indicating that infants from more disadvantaged households tend to score lower on the cognitive subscale at age 2. There was also no association between maternal education and infant gender, but higher maternal education was associated with higher SES. Finally, there was an additional negative association between maternal education and parity, and there was no evident trend for the presence of an ACG to be associated with chronic or antenatal depression.

Following a similar pattern, when bivariate associations were calculated separately for boys and girls, chronic depression showed fewer significant associations with other variables for boys than was evident in the early postnatal sample. However, other than the absence of an association between cognitive development and the presence of an ACG, associations with cognitive development were similar between samples for boys. SES also appeared to be more important in this sample, as it was negatively associated with both maternal education and the presence of an ACG. As it is reverse coded this suggests that higher SES was related to higher education and an increased likelihood of the presence of an ACG. Patterns of association for girls were identical between samples, other than the absence of an association between cognitive development and parity in the chronic sample.

Table 3.3.2.2a: Bivariate associations between key study variables in chronic exposure sample

	Language	Chronic Depression	AND[b]	SND[c]	Maternal Age	Education	Parity	SES	ACG[d]	Gender
Cognitive	**.55****	-.04	-.06	.02	.07	.00	-.03	**-.11***	.02	.04
Language		-.06	-.05	.07	.02	.04	**-.13***	-.07	.04	**.14****
Chronic			**.23****	-.05	-.07	-.05	.03	.01	0.08	-.08
AND				.00	-.09	.02	-.10	.09	-.12	.10
SND					.10	.10	-.02	-.01	.04	.02
Mat Age						.00	**.41****	-.06	-.03	.04
Education							**-.16***	**-.24****	.08	-.15†
Parity								.14†	-.10	.05
SES									-.16†	.08
ACG										-.08

Note – [a] Postnatal Depression; [b] Antenatal Depression; [c] Sensitivity to Non-Distress; [d] Alternate Caregiver; * p < .05; ** p < .001; Tetrachoric correlations reported for associations between categorical predictors.

Table 3.3.2.2b: Bivariate associations between key variables in chronic exposure sample separately for boys and girls. Boys (blue) in the top diagonal, girls (green) in the bottom diagonal.

	Cognitive	Language	Chronic Depression	AND[b]	SND[c]	Maternal Age	Education	Parity	SES	ACG[d]
Cognitive		**.56****	-.03	-.01	.09	**.16***	.00	.03	**-.22****	.05
Language	**.53****		-.13	-.11	**.18***	.09	-.01	.00	-.15	.00
Chronic Depression	-.04	-.05		**.19****	-.02	-.07	-.11	.00	.05	-.10
AND	-.12	-.06	**.26****		.01	**-.19****	-.02	-.21†	.12	-.22†
SND	-.05	-.04	-.08	.01		.13	**.18****	.02	-.04	.06
Mat Age	-.03	-.01	-.06	.01	.07		.01	**.46****	-.07	-.04
Education	.01	.06	.00	.01	.01	.01		-.20†	**-.25***	.10
Parity	-.11	**-.20****	.06	.02	-.07	**.36****	-.11		.14	0.18
SES	.03	.00	-.02	.07	.02	-.05	-.21†	.15		**-.28***
ACG	.11	.01	-.07	-.03	.02	-.02	.04	-.01	-.04	

Note – [a] Postnatal Depression; [b] Antenatal Depression; [c] Sensitivity to Non-Distress; [d] Alternate Caregiver; * p < .05; ** p < .001; Tetrachoric correlations reported for associations between categorical predictors.

3.3.4 Multivariate analysis

The effects of early postnatal depression and chronic depression on each outcome were investigated independently. Separate linear regression analyses were performed investigating each combination of exposure (postnatal or chronic) and outcome (cognitive or language) first in the whole sample and then by sex. Both mediation and moderation analyses were planned in line with study hypotheses, but the lack of main effect precluded testing the mediational models. Therefore, only main and interactive effects were tested. Results are presented below.

3.3.4.1 Early postnatal depression and cognitive development (Table 3.3.4.1)

Full Timing Sample. In the first step, postnatal depression did not significantly predict cognitive development. Postnatal depression remained non-significant after adjustment for covariates. None of the included covariates were significant and the overall model was non-significant (p = .549). In the third step, the main effect of sensitivity to non-distress was added. This was non-significant and the block was not a significant improvement to the model (p = .494). In the final block, the interaction term between postnatal depression and SND was added. The interaction term was non-significant and the step did not represent improvement to the model (p = 813). Overall, the model was non-significant and accounted for none of the variance in cognitive development (F(10)=.74, p = .688).

Boys. Postnatal depression did not significantly predict cognitive development in boys, before or after adjusting for covariates. However, both lower SES and the presence of an ACG predicted lower cognitive scores. Step 2 represented a significant overall model (p = .014). The main effect of SND was added in the third step, was non-significant and did not improve the overall model (p = .261). In the final step, the interaction term between postnatal depression and SND was included. The interaction term was non-significant and did not improve the overall model (p = .792). Overall, the model was significant and accounted for 6.1% of variance in cognitive development (F(9)=2.17, p = .027).

Girls. Postnatal depression was not a significant predictor of cognitive development in girls, before or after adjusting for covariates. However, in the second step, multiparity predicted poorer cognitive development. Further, in contrast to the findings for boys, there was a marginal, positive effect of the presence of an ACG. The overall block failed to reach the conventional level of statistical significance (p = .098). In the third block, the main effect of SND was entered and was non-significant. This block did not significantly improve the overall model (p = .909). In the final block, the interaction term between postnatal

depression and SND was entered. This effect was non-significant and the block did not improve the overall model (p = .833). The final model was non-significant and accounted for 2.2% of variance in cognitive development in girls (F(9)=1.36, p = .211).

Table 3.3.4.1: Summary of linear regression models predicting cognitive development from early postnatal depression, sensitivity to non-distress, and other covariates

Cognitive Development	Full Sample		Boys		Girls	
	***b* (95% CI)**	***p***	***b* (95% CI)**	***p***	***b* (95% CI)**	***p***
Step 1						
8-week Maternal Depression	-1.04 (-3.30, 1.23)	0.368	-.74 (-3.89, 2.42)	.644	-1.42 (-4.72, 1.89)	0.398
Model Statistics	*F (1, 307) = .81, p = .368, $R^{2-=}$ .00*		*F (1, 163) = .21, p = .644, R^2 = .00*		*F (1, 142) = .72, p = .398, R^2 = 00*	
Step 2						
8-week Maternal Depression	-0.52 (-2.97, 1.94)	0.679	-1.05 (-4.31, 2.22)	0.526	0.21 (-3.40, 3.82)	0.908
Antenatal Depression	-1.04 (-3.04, 0.96)	0.306	0.20 (-2.56, 2.95)	0.888	-1.94 (-4.78, 0.91)	0.180
Maternal Education	-0.86 (-2.95, 1.23)	0.418	-1.52 (-4.29, 1.25)	0.281	-0.13 (-3.16, 2.91)	0.934
Maternal Age	0.12 (-0.16, 0.40)	0.407	0.34 (-0.07, 0.75)	0.103	-0.05 (-0.43, 0.32)	0.780
SES	-1.51 (-3.51, 0.69)	0.121	-3.76 (-6.41, 1.12)	0.006*	0.67 (-1.99, 3.33)	0.620
Parity	-1.41 (-3.51, 0.69)	0.188	-0.31 (-3.24, 2.61)	0.833	-3.11 (-6.00, -0.22)	0.035*
Alternate Caregiver	-0.62 (-2.99, 1.74)	0.603	-4.53 (-8.05, 1.01)	0.012*	2.79 (-0.25, 5.83)	0.072+
Child Gender	-0.01 (-1.90, 1.88)	0.992	N/A	N/A	N/A	N/A
Model Statistics	*F (8, 300) = .86, p = .549, R^2 = .00*		*F (7, 157) = 2.62, p = .014, R^2 = .07*		*F (7, 136) = 1.77, p = .098, R^2 = .04*	
Step 3						
Sensitivity to Non-Distress	0.34 (-0.64, 1.32)	0.494	0.85 (-0.64, 2.33)	0.261	-0.07 (-1.33, 1.19)	0.909
Model Change Statistics	*F (1, 299) = .47, p = .494, R^2 = .00*		*F (1, 156) = 1.28, p = .261, R^2 = .07*		*F (1, 135) = .01, p = .909, R^2 = .03*	
Step 4						
Maternal Depression * Sensitivity to Non-Distress	-0.29 (-2.70, 2.12)	0.813	-0.48 (-4.04, 3.09)	0.792	-0.34 (-3.50, 2.82)	0.833
Model Change Statistics	*F (1, 298) = .06, p = .813, R^2 = .00*		*F (1, 155) = .07, p = .792, R^2 = .06*		*F (1, 134) = .04, p = .833, R^2 = .02*	

3.3.4.2 Early Postnatal depression and language development (Table 3.3.4.2).

Full Timing Sample. In the first step, postnatal depression was a significant, independent predictor of language development (*B* = -3.37, *p* = .037). However, this effect became non-significant after adjusting for covariates. Infants whose mothers were multiparous had significantly lower language scores, as did male infants. Overall, the second block was significant (*p* = .049). In the third step, the main effect of sensitivity to non-distress was added. This was non-significant and the block was not a significant improvement to the model (*p* = .413). In the final block, the interaction term between postnatal depression and SND was added. The interaction term was non-significant and the step did not represent improvement to the model (*p* = .796). Overall, the model was marginally non-significant and accounted for 2.1% of variance in language development (F(10)=1.65, *p* = .092).

Boys. In the first step there was a marginal, but non-significant negative effect of postnatal depression on language development (*B* = -3.58, p = .077). This effect became non-significant after adjusting for covariates in the second step. None of the covariates were significant predictors of development and the overall block was non-significant (*p* = .333). The main effect of SND was added in the third step and it was found that higher sensitivity predicted higher language scores. The improvement to the overall model was significant (*p* = .032). In the final step, the interaction term between postnatal depression and SND was included. The interaction term was non-significant and did not improve the overall model (*p* = .517). The final model was non-significant and accounted for 2.6% of variance in language development (F(9)=1.48, *p* = .159).

Girls. Postnatal depression did not make a significant contribution to language development in girls, before or after controlling for covariates. In the second block, infants whose mothers were multiparous scored significantly lower on the language scale. The overall block was non-significant (*p* = .115). In the third block, the main effect of SND was entered and was non-significant. The block did not significantly improve the overall model (*p* = .514). In the final block, the interaction term between postnatal depression and SND was entered. This effect was non-significant and the block did not improve overall model fit (*p* = .660). The final model was non-significant accounted for 2.3% of variance in language development in girls (F(9)=1.38, *p* = .205).

Table 3.3.4.2: Summary of linear regression models predicting language development from eary postnatal depression, sensitivity to non-distress, and other covariates

Language Development	Full Sample		Boys		Girls	
	***b* (95% CI)**	***p***	***b* (95% CI)**	***p***	***b* (95% CI)**	***p***
Step 1						
8-week Maternal Depression	-3.37* (-6.54, -0.20)	.037	-3.58 (-7.55, 0.39)	.077	-2.78 (-7.92, 2.35)	0.285
Model Statistics	*F (1, 307) = 4.37, p = 0.037, R^2 = .01*		*F (1, 163) = 3.16, p = .077, R^2 = .01*		*F (1, 143) = 1.15, p = .285, R^2 = .00*	
Step 2						
8-week Maternal Depression	-2.07 (-5.47, 1.33)	0.232	-2.96 (-7.24, 1.31)	0.173	-0.26 (-5.89, 5.36)	0.927
Antenatal Depression	-1.75 (-4.52, 1.02)	0.216	-1.59 (-5.19, 2.01)	0.385	-1.81 (-6.23, 2.62)	0.421
Maternal Education	-0.37 (-3.27, 2.53)	0.802	-1.13 (-4.76, 2.50)	0.539	0.64 (-4.09, 5.39)	0.789
Maternal Age	0.28 (-0.12, 0.67)	0.166	0.21 (-0.33, 0.75)	0.436	0.36 (-0.23, 0.95)	0.224
SES	-0.67 (-3.32, 1.99)	0.622	-2.82 (-6.28, 0.65)	0.110	1.87 (-2.27, 6.01)	0.373
Parity	-3.58 (-6.50, -0.67)	0.016*	-0.72 (-4.54, 3.11)	0.712	-7.14 (-11.64, -2.64)	0.002*
Alternate Caregiver	-0.51 (-3.79, 2.76)	0.758	-1.38 (-5.99, 3.23)	0.555	0.20 (-4.53, 4.93)	0.934
Child Gender	2.53 (-0.10, 5.15)	0.059*	N/A	N/A	N/A	N/A
Model Statistics	*F (8, 300) = 1.98, p = .049, R^2 = .03*		*F (7, 157) = 1.15, p = .333, R^2 = .01*		*F (7, 136) = 1.70, p = .115, R^2 = .03*	
Step 3						
Sensitivity to Non-Distress	0.57 (-0.79, 1.93)	0.413	2.11 (0.19, 4.03)	0.032*	-0.65 (-2.60, 1.31)	0.514
Model Change Statistics	*F (1, 299) = .67, p = .413, R^2 = .02*		*F (1, 156) = 4.69, p = .032, R^2 = .03*		*F (1, 135) = .43, p = .514, R^2 = .03*	
Step 4						
Maternal Depression * Sensitivity to Non-Distress	0.44 (-2.90, 3.78)	0.796	1.52 (-3.10, 6.13)	0.517	-1.09 (-2.62, 1.75)	0.660
Model Change Statistics	*F (1, 298) = .07, p = .796, R^2 = .02*		*F (1, 155) = .42, p = .517, R^2 = .03*		*F (1, 134) = .19, p = .660, R^2 = .02*	

3.4.3.3 Chronic depression & cognitive development (Table 3.4.3.3).

Full Chronic Sample. Chronic depression did not significantly predict cognitive development, independently or after adjustment for covariates. In the second step there was a marginal effect of SES, with lower SES predicting poorer cognitive scores, but none of the other covariates were significant and the overall model was non-significant (p = .309). In the third step, the main effect of sensitivity to non-distress was added. This was non-significant and the block was not a significant improvement to the model (p = .776). In the final block, the interaction term between chronic depression and SND was added. The interaction term was non-significant and the step did not represent improvement to the model (p = 884). Overall, the model was non-significant and accounted for none of the variance in cognitive development (F(10)=.69 p = .486).

Boys. Chronic depression did not significantly predict cognitive development in boys, independently or after adjustment for covariates.. In the second step there was a significant and negative effect of lower SES and a marginal positive effect of maternal age. Unlike in the postnatal model, the presence of an ACG was non-significant but the overall block was still significant (p = .029). The main effect of SND was added in the third step, was non-significant, and did not improve the overall model (p = .258). In the final step, the interaction term between chronic depression and SND was included. The interaction term was non-significant and did not improve the overall model (p = .182). Overall, the model was significant and accounted for 6.1% of variance in cognitive development (F(9)=2.17, p = .028).

Girls. Chronic depression did not significantly predict cognitive development in girls, independently or after adjustment for covariates. Unlike the postnatal model, in the second step, multiparity did not predict poorer cognitive development. However, although non-significant, the chronic model shows a similar pattern of findings regarding the presence of an ACG, which shows a negative trend for boys and a positive trend for girls. The overall block was non-significant (p = .323). In the third block, the main effect of SND was entered and was non-significant. The block did not significantly improve the overall model (p = .357). In the final block, the interaction term between postnatal depression and SND was entered. This effect was non-significant and the block did not improve overall model fit (p = .173). The final model was non-significant and accounted for 2.2% of variance in cognitive development in girls (F(9)=1.36, p = .288).

Table 3.4.3.3: Summary of linear regression models predicting cognitive development from chronic depression, sensitivity to non-distress, and other covariates

Cognitive Development	Full Sample		Boys		Girls	
	***b* (95% CI)**	***p***	***b* (95% CI)**	***p***	***b* (95% CI)**	***p***
Step 1						
Chronic Maternal Depression	-.46 (-1.47, 0.55)	0.372	.00 (-1.44, 1.44)	0.999	-.97 (-2.40, 0.45)	.179
Model Statistics	*F (1, 393) = .79, p = .372, R^2 = 00*		*F (1, 203) = .00, p = 0.999, R^2 = .00*		*F (1, 188) = 1.82, p = .179, R^2 = .00*	
Step 2						
Chronic Maternal Depression	-0.20 (-1.27, 0.86)	0.706	0.16 (-1.31, 1.63)	0.831	-0.49 (-2.01, 1.03)	0.528
Antenatal Depression	-0.76 (-2.50, 0.98)	0.390	0.29 (-2.21, 2.78)	0.817	-1.67 (-4.07, 0.73)	0.172
Maternal Education	-0.35 (-2.14, 1.43)	0.698	-0.59 (-3.10, 1.92)	0.642	0.01 (-2.49, 2.50)	0.997
Maternal Age	0.20 (-0.05, 0.44)	0.117	0.38 (-0.01, 0.77)	0.053+	0.05 (-0.26, 0.36)	0.729
SES	-1.60 (-3.27, 0.08)	0.062+	-3.90 (-6.33, -1.46)	0.002*	0.71 (-1.55, 2.98)	0.535
Parity	-1.01 (-2.82, 0.79)	0.270	-0.41 (-3.09, 2.27)	0.763	-1.88 (-4.27, 0.51)	0.123
Alternate Caregiver	0.14 (-1.18, 2.17)	0.890	-1.77 (-4.84, 1.30)	0.258	1.87 (-0.76, 4.50)	0.162
Child Gender	0.60 (-1.04, 2.25)	0.472	N/A	N/A	N/A	N/A
Model Statistics	*F (8, 386) = 1.18, p = .309, R^2 = .00*		*F (7, 197) = 2.29, p = .029, R^2 = .04*		*F (7, 182) = 1.17, p = .323, R^2 = .01*	
Step 3						
Sensitivity to Non-Distress	0.12 (-0.71, 0.95)	0.776	0.74 (-0.55, 2.03)	0.258	-0.49 (-1.54, 0.56)	0.357
Model Change Statistics	*F (1, 385) = .08, p = .776, R^2 = .00*		*F (1, 196) = 1.28, p = .258, R^2 = .04*		*F (1, 181) = .85, p = .357, R^2 = .01*	
Step 4						
Maternal Depression * Sensitivity to Non-Distress	-0.08 (-1.12, 0.96)	0.884	-1.06 (-2.61, 0.50)	0.182	0.94 (-0.41, 2.29)	0.173
Model Change Statistics	*F (1, 384) = .02, p = .884, R^2 = .00*		*F (1, 195) = 1.79, p = .182, R^2 = .05*		*F (1, 180) = 1.87, p = .173, R^2 = .01*	

3.4.3.4 Chronic depression and language development (Table 3.4.3.4)

Full Chronic Sample. Chronic depression did not significantly predict language development, independently or after adjustment for covariates. In the second step, infants whose mothers were multiparous had significantly lower language scores, as did male infants. Overall, the block was significant (p = .006). In the third step, the main effect of sensitivity to non-distress was added. This was non-significant and the block was not a significant improvement to the model (p = .324). In the final block, the interaction term between chronic depression and SND was added. The interaction term was non-significant and the step did not represent improvement to the model (p = .821). Overall, the model was significant and accounted for 3.1% of variance in language development (F(10)=2.23, p = .013).

Boys. Chronic depression did not significantly predict cognitive development in boys, independently or after adjustment for covariates. In the second step, there was a marginal and negative effect of lower SES, but none of the covariates were significant predictors of development and the overall block was non-significant (p = .357). The main effect of SND was added in the third step and, consistent with the postnatal model, it was found that higher sensitivity predicted higher language scores. The improvement to the overall model was significant (p = .015). In the final step, the interaction term between chronic depression and SND was included. The interaction term was non-significant and did not improve the overall model (p = .603). The final model was non-significant and accounted for 2.5% of variance in language development (F(9)=1.58, p = .125).

Girls. Chronic depression did not significantly predict cognitive development in girls, independently or after adjustment for covariates. In the second step, infants whose mothers were multiparous scored significantly lower on the language scale, indicating that the effect found in the full sample is unique to girls. The overall block was not significant (p = .124). In the third block, the main effect of SND was entered and was non-significant. The block did not significantly improve the overall model (p = .306). In the final block, the interaction term between chronic depression and SND was entered. This effect was non-significant and the block did not improve overall model fit (p = .536). The final model was non-significant and accounted for 2.0% of variance in language development in girls (F(9)=1.44, p = .174).

Table 3.4.3.4: Summary of linear regression models predicting language development from chronic depression, sensitivity to non-distress, and other covariates

Language Development	Full Sample		Boys		Girls	
	***b* (95% CI)**	***p***	***b* (95% CI)**	***p***	***b* (95% CI)**	***p***
Step 1						
Chronic Maternal Depression	-.81 (-2.23, 0.61)	.263	-.06 (-1.96, 1.83)	0.948	-1.41 (-3.55, 0.74)	0.197
Model Statistics	*F (1, 393) = 1.26, p = .263, R^2 = .00*		*F (1, 203) = .00, p = .948, R^2 = .00*		*F (1, 188) = 1.68, p = .197, R^2 = .00*	
Step 2						
Chronic Maternal Depression	-0.19 (-1.66, 1.29)	0.804	0.43 (-1.53, 2.39)	0.667	-0.89 (-3.15, 1.37)	0.437
Antenatal Depression	-1.28 (-3.69, 1.13)	0.297	-1.24 -(4.59, 2.11)	0.467	-0.92 (-4.49, 2.65)	0.612
Maternal Education	0.65 (-1.82, 3.12)	0.605	0.34 (-3.01, 3.70)	0.840	1.16 (-2.55, 4.86)	0.539
Maternal Age	0.27 (-0.07, 0.62)	0.113	0.22 (-0.30, 0.73)	0.409	0.30 (-0.16, 0.76)	0.194
SES	-1.24 (-3.57, 1.08)	0.293	-3.17 (-6.42, .09)	0.057+	0.79 (-2.58, 4.16)	0.645
Parity	-3.67 (-6.17, -1.17)	0.004*	-1.93 (-5.51, 1.65)	0.289	-5.39 (-8.95, -1.83)	0.003*
Alternate Caregiver	0.74 (-2.06, 3.55)	0.602	1.28 (-2.82, 5.39)	0.538	0.19 (-4.43, 2.71)	0.926
Child Gender	3.30 (1.02, 5.59)	0.005*	N/A	N/A	N/A	N/A
Model Statistics	*F (8, 386) = 2.73, p = .006, R^2 = .03*		*F (7, 197) = 1.11, p = .357, R^2 = .00*		*F (7, 182) = 1.65, p = .124, R^2 = .02*	
Step 3						
Sensitivity to Non-Distress	0.58 (-0.57, 1.72)	0.324	2.11 (0.41, 3.82)	0.015*	-0.81 (-2.38, 0.75)	0.306
Model Change Statistics	*F (1, 385) = .98, p = .324, R^2 = .03*		*F (1, 196) = 5.97, p = .015, R^2 = .03*		*F (1, 181) = 1.06, p = .306, R^2 = .02*	
Step 4						
Maternal Depression * Sensitivity to Non-Distress	0.16 (-0.57, 1.73)	0.821	-0.55 (-2.61, 1.52)	0.603	0.63 (-1.38, 2.65)	0.536
Model Change Statistics	*F (1, 384) = .05, p = .821, R^2 = .03*		*F (1, 195) = .271, p = .603, R^2 = .03*		*F (1, 180) = .385, p = .536, R^2 = .02*	

3.5 Discussion

3.5.1 Summary of main findings

The present study found a significant, unadjusted, and small negative effect of early postnatal depression on language development (*d* = .29, .01-.56). This effect became marginal but remained at a similar magnitude when examined in boys only and was smaller and non-significant in girls. However, after adjusting the effects of covariates, neither early postnatal nor chronic maternal depression significantly predicted cognitive or language development. Maternal sensitivity did predict improved language outcomes in boys only but was not associated with cognitive outcomes. There was also no evidence that maternal sensitivity moderated the relationship between maternal depression and cognitive or language development, either in the full sample or the sex-specific analyses. Some interesting sex-specific findings did emerge regarding the covariates. Lower SES predicted poorer cognitive outcomes in boys only, while multiparity predicted poorer cognitive and language outcomes in girls only. Additionally, the presence of an alternate caregiver predicted significantly worse cognitive outcomes for boys, while for girls the association was positive, albeit marginally non-significant.

3.5.2 Comparison of findings regarding early postnatal depression with HIC and LMIC literature.

The systematic review presented in chapter 2 of this thesis revealed inconsistent findings regarding the direct relationship between maternal postnatal depression and infant cognitive development. Out of the nine studies reviewed, only four found a significant effect of depression, and one of these was cross-sectional. Thus, the current finding from the multivariate analyses that early postnatal depression did not significantly predict cognitive development at 2 years is not entirely unexpected. However, the finding that chronic maternal depression over the first 2 years is similarly unrelated to cognitive outcomes was more surprising. Although there are far fewer studies that have explored chronic exposure to depression in LMIC settings, findings from HIC settings are indicative of far more consistent negative effect.

Interestingly, the finding of a small unadjusted effect of early postnatal depression on language subscale scores at 2 years (*d* = .29) is remarkably similar to the unadjusted effect of 0.28 reported by a recent meta-analysis examining the effect of maternal depression on cognitive development in a mix of HIC and LMIC settings (Liu et al., 2017), although it should be noted that the current findings regarding the cognitive subscale did not show the same similarities. Additionally, unlike in the present study, after restricting the meta-

analysis to a subset of studies that assessed the impact of maternal depression at 6-8 weeks postpartum using multivariate analyses, the adjusted effect remained significant, indicating a stronger effect of postnatal depression in these studies than in studies which assessed depression at a later developmental timepoint.

There are a number of methodological and sociocultural factors that may influence comparisons between the present and previous studies which will now be discussed.

3.5.2.1 Methodological comparison of studies investigating role of early postnatal depression.

The prevalence of depression in the current sample was highest at 8 weeks (19.3%) and decreased slightly but then remained fairly uniform from 6-24 months (14.4-15.8%). These rates are not dissimilar to those found in a recent meta-analysis of Indian studies (19%, Upadhyay et al., 2017), indicating that the present sample was not atypical of this setting.

There are only two other empirical studies of a similar nature that have carried out in India. Thomas et al. (2020) investigated the effects of antenatal depression in an urban sample from Bangalore. Although the exposure is not directly comparable since it was antenatal, their sample was also drawn from urban Bangalore and the results were remarkably similar to the present study. They reported finding a weak univariate association between antenatal depression and language development and that this became non-significant after adjusting for covariates. Thus, it is possible that factors specific to urban populations in India may be having a greater influence on cognitive development than perinatal depression. In the other study, Patel, DeSouza and Rodrigues (2003) recruited a community sample from a mix of urban and rural areas in Goa and found that postnatal depression assessed at 6 weeks predicted poorer cognitive development in infants at 6 months, after controlling for a number of covariates. In addition to sociodemographic differences between rural Goa and urban Bangalore, a number of key differences may explain the disparity in findings. Firstly, Patel and colleagues assessed cognitive development at 6 months compared to 24 months in the present study. It is possible that even if development is particularly sensitive to external factors during the postnatal period, this effect may be short lived. Additionally, it is difficult to reliably measure cognitive development at this early age, and any evaluation at this point may not reflect developmental changes later in life (Quevedo et al., 2012). This may be because the developing brain is sensitive to its environment to the degree that development fluctuates significantly over short periods of time and that it is only later in infancy that it settles into a more consistent pattern of functioning.

In Barbados, Galler, Harrison, Ramsey, Forde and Butler (2000) found a significant effect of maternal depression at 7 weeks on cognitive development at 3 months. At 6 months, although a similar trend was present, it was no longer significant. Another study in Brazil did find a significant effect of depression on language development at 12 months, indicating a more long-term effect (Quevedo et al., 2012). However, mothers in this study received a clinical diagnosis of depression rather self-reporting symptoms via a screening questionnaire, and the effect was strongest when depression was present at both 2 months and 12 months. In comparison, each of the LMIC studies that did not find a significant effect of early postnatal depression assessed depression at a single time-point using a screening questionnaire and, with the exception of Tran et al., (2012), all assessed cognition at 12 months or later. It may be that mild to moderate postnatal depression can have a short-term impact on development, but long-term effects are restricted to children of mothers with more severe and enduring clinically diagnosed depression. Therefore, in the current study it is possible that any short-term effect of early PND may have dissipated by 24 months and so was undetected.

There are other differences between the current study and those that found a significant effect which may also be responsible for contrasting findings. While the current study used the BSID-III, Patel et al., (2003) used the Developmental Assessment Scales for Indian Infants (DASII). As the BSID-III has not been validated in India, it is possible that the DASII represents a more accurate assessment of cognitive development in this context. However, the BSID-III is has been widely used in India and other LMIC settings, and other studies have found significant impairments using this tool, including Quevedo et al. (2012). Although the current study did not look explicitly at rates of developmental delay, it is worth noting that the rates of children identified to be at risk for mild cognitive (6.5%) or language (10.6%) delay in this study were substantially lower than the rates found by Patel and colleagues (31.4%). This difference may be an artefact of the different tools used, but it could also indicate that the Patel sample was generally more high-risk. Socioeconomic status or income were not reported so direct comparison of risk is not possible. However, assessments were performed in a nested subsample of depressed mothers matched with non-depressed controls, resulting a far higher proportion of depressed mothers than is typically found in community samples. Therefore it is likely that the level of risk in the sample was higher. This approach also generates more extreme comparisons which in turn can lead to larger effect sizes than found in traditional community samples (Bernard et al., 2018).

A further point of difference is that the present study controlled for the effects of antenatal depression. The lack of control for antenatal depression was highlighted in the systematic review as one of the key limitations of the LMIC literature and has important implications. It is a critical issue because the presence of antenatal depression may confound the effect of postnatal depression (Osborne et al., 2018; Rahman & Creed, 2007), and so the lack of control severely undermines confidence in any findings. Thus, it is possible that the effects observed in those studies are at least partially due to antenatal, rather than postnatal, effects. This point is emphasised by the fact that the only study in the review that did not find a significant effect at 6 months also was the only study to control for the effects of antenatal depression (Tran et al., 2012). However, in contrast to Tran et al. (2012) who reported that antenatal depression was associated with poorer cognitive development, the present study did not find evidence of a significant association between antenatal depression and cognitive development, either in the bivariate or multivariate analyses. Thus, it is unlikely that controlling for antenatal depression had the same effect in the current study.

There is a similar pattern in HIC research, although not quite as clearly delineated. Three out of four prospective studies found significant effects of postnatal depression on cognitive outcomes in the first 12 months. Two of these assessed depression using a clinical interview, with one finding an effect at 12 months (Perra et al., 2015) while the other reported an effect at 4 months that had dissipated by 13 months (Smith-Nielsen et al., 2016). A third study found an effect of self-reported depression on cognitive development at 12 months in very high-risk sample (Lyons-Ruth et al., 1986). None controlled for the effects of antenatal depression, although Perra et al. (2015) did account for past history of depressive illness. Findings regarding the long-term effects of self-reported depression were less consistent, although more studies reported a lack of significant association beyond 12 months (Kiernen & Huerta, 2012; Piteo, Yelland, & Makrides, 2012; Stein et al., 2008) than reported a significant association (Kawai et al 2017; Koutra et al., 2013). Both studies which grouped mothers using a clinical diagnosis found a significant long-term effect (Conroy et al., 2012; Murray, 1992).

3.5.3 Methodological comparison of studies investigating the role of chronic depression.

Very few studies have explored the impact of chronic exposure to depression in LMICs and so there is very little the present results can be compared to in this context. Quevedo et al.'s (2012) study has already been discussed and the key difference has been highlighted in terms of the inclusion of a clinical diagnosis. However, whether depression being present at

2 months and 12 months can be considered chronic is questionable. In the only other study, depression was assessed using the EPDS at 2 weeks, 6 months and 18 months (Garman et al., 2019). Symptoms were modelled as trajectories, including early postpartum (2 weeks), late postpartum (18 months) and chronic high. The authors did not find any significant effect of chronic depression on cognitive outcomes assessed by BSID-II at 18 months or an executive function battery at 36 months. Although the use of trajectories differs from the present use of a factor score, there are broad similarities in the number and timing of depression assessments, as well as the later age at which cognition was assessed.

Conversely, almost all HIC studies that investigated chronic symptoms of depression found a significant deleterious impact on cognitive development. These findings stand in stark contrast to the present study but also contribute to an increasingly consistent pattern. Several key methodological features stand out that are likely to be driving disparities. Firstly, the use of a clinical diagnostic measure of depression was common to almost all studies (Azak, 2012; Cornish et al., 2005; Milgrom, Westley, & Gemmill, 2004; Park, Brain, Grunau, Diamond, & Oberlander, 2018). As already discussed, this approach is likely to yield groups of mothers with more severe depression than groups demarcated by a cut-off score on a screening questionnaire. In addition, two of these studies recruited parallel groups of clinic referred and healthy control mothers (Azak, 2012; Milgrom et al., 2004), an approach likely to yield larger effect sizes (Bernard, et al., 2018).

The frequency of depression assessments and the conceptualisation of chronicity are also important features of the HIC studies. Differences in approach between the present study and the HIC studies are not as pronounced in this area but combined with the use of diagnostic measures speak to the resolution required to detect significant effects. For example, there are two studies which used screening questionnaires exclusively. In one, the CES-D was administered at five points in the first 3 years (NICHD Early Child Care Research Network (ECRN), 1999). Mothers were only included in the chronic group if they scored above the cut-off score on at least four assessments. In the other study, the EPDS was administered at six time-points across the first 2 years, but chronicity was conceptualised as the mean EPDS score across follow-up (Sutter-Dalley, et al., 2011). The former found a significant effect of chronicity, while the latter did not, indicating that the use of a clinically relevant cut-off created a more severely group of depressed mothers than using a mean score. This example illustrates how sensitive results can be to methodological approaches. Although the present approach of creating a factor score from EPDS assessments at four time-points is sufficiently robust, it is less strict than the NICHD ECRN approach, which had

more assessment points and very demanding criteria for inclusion in the chronic group. These factors, in addition to the absence of a clinical diagnosis and the distinct cultural context, may be responsible for the disparity between the findings reported here and those reported by the NICHD ECRN (1999).

3.5.4 When are differences in cognitive development predicted by maternal depression?

By drawing together the findings from HIC and LMIC studies discussed above with the findings from the present study, a consistent picture can be seen to emerge. Clearly differences are most often detected when depression is diagnosed using a clinical diagnostic measure rather than a screening questionnaire. Chronic exposure does exacerbate the effects of maternal depression on cognitive development, but generally only when that depression is at a clinical level. This is illustrated by findings from the present study and by other studies in both LMIC and HIC settings (Garman et al., 2019; Sutter-Dalley et al., 2011) which found no effect of chronic depression assessed using screening questionnaires. The only apparent exception to this is the NICHD ECRN (1999) study, whose intensive approach both to assessment frequency and conceptualisation of chronicity somewhat sets them apart. Mild-moderate early postnatal or chronic depression, then, does not appear to impair infant cognitive development. While this is good news in that it means these children may not be at a developmental disadvantage, there should also be some caution. Screening tools, such as the Edinburgh Postnatal Depression Scale, are very common in both HIC and LMIC research, mainly due to their advantages in terms of ease and speed of administration, particularly in larger samples. It is, therefore, possible that this reliance on screening tools is precluding investigation of the true effects of more severe depression and as a result, potentially harmful effects of depression in LMIC settings are being masked.

3.5.5 Comparison of findings regarding the role of caregiving with HIC and LMIC literature

There was no evidence of a moderating effect of postnatal depression by maternal sensitivity, on cognitive or language development outcomes either in the full sample or in sex-specific analyses. However, postnatal depression was weakly but significantly associated with lower levels of maternal sensitivity in bivariate analyses. In multivariate analyses, there was evidence of a sex-specific main effect of sensitivity, with higher sensitivity predicting better language development in boys only. This finding fits with previous work in HIC settings indicating that boys may be particularly sensitive to the early caregiving environment (Azak, 2012; Bridgett et al., 2009; Murray et al., 1996, 2010) but the small magnitude of effects also need to be set in the context of this being a community

sample. It is possible that the level of depression detected by the EPDS in the current sample simply was not severe enough to compromise caregiving substantially. However, it is also possible that the relationship between maternal depression and caregiving functions differently in this context. Each possibility will be discussed in the following sections.

There have been very few studies with this focus in LMIC settings. Black et al. (2007) did find that that caregiving, involving increased sensitivity and infant stimulation from the mother, was significantly and negatively associated with maternal depression, and that it mediated the effect of a *depression X temperament* interaction on cognitive outcomes. However, this was a cross-sectional effect and the authors did not explore whether caregiving had a role in the direct relationship between depression and later development. Comparisons can also be drawn between the present study and NICHD ECRN (1999). In that study, maternal depression predicted significant impairments in sensitivity, and sensitivity itself was a significant predictor of child cognitive outcomes in boys and girls. Sex-specific effects were not explored, but further analysis revealed significant mediation and moderation effects. In discussing the differences between these and the present findings some important distinctions need to be made. Firstly, in the NICHD ECRN study caregiving was assessed using the same play-based procedure and coding scheme as the current study, but the final variable included in the moderation/mediation analyses was created by taking the mean score from four time-points across the first 3 years, as opposed to a single time point in this study. Thus, it is possible that long-term impairments in maternal sensitivity are required to evoke changes in developmental outcomes. They also observed a significant decline in sensitivity in depressed mothers across the second year and then some recovery in the third year. It is possible that the current study, which assessed caregiving at 6-12 months, did not capture the effects on sensitivity at a potentially difficult point of development, where infants exhibit behaviours which make interactions more challenging for depressed mothers. Further, the current study only used the 'sensitivity to non-distress' subscale, as it considered to be the most relevant to cognitive development, but the NICHD study created a composite variable from 'sensitivity to non-distress', 'intrusiveness', and 'positive regard'. Similarly, Gueron-Sela et al. (2018) also found a mediating effect of caregiving, albeit at a later age and after using explanatory factor analysis to produce two factors from the individual NICHD scales. Milgrom et al. (2004) also reported that maternal responsiveness at 6 months acted as a mediator between postnatal depression and child IQ scores at 42 months, but not for cognitive/language profile scores. As mentioned earlier, mothers had received a diagnosis of clinical depression, and so the

impact of depression on both caregiving and development is likely to have been stronger than in the present study.

The current study also explored gender-specific effects of both maternal depression and sensitivity. Effects were explored separately in each gender, rather than by exploring a three-way interaction, because the bivariate associations indicated a stronger effect in boys, rather than an opposite effect between genders, and also because of concerns regarding statistical power. In terms of sex-differences, there is strong evidence of a unique risk to boys from depression (Azak, 2012; Hay et al., 2001; Milgrom et al., 2004; Sharp et al., 1996) and that these differences are exacerbated in the context of compromised caregiving, with depressed mothers being less sensitive and expressing fewer affirmations toward male infants (Murray et al., 1996) and showing less overall infant focused speech but a higher proportion of critical and hostile speech toward male infants (Murray et al., 1996, 2010). This vulnerability is thought to be driven by male infants being less able self-regulate their emotional and attentional states, thereby making them more reliant on facilitation and scaffolding of an emotionally healthy caregiver (Murray & Cooper, 1997). It is also possible that male infants, being higher in negative affect and less able to self-regulate that affect, are more difficult to parent, thereby eliciting a more negative style of parenting from mothers, particularly when their resources are already limited by the presence of depressive symptoms (Crockenberg & Leerkes, 2003). Although there was no evidence of a sex-specific moderation effect in the present study, there was a significant and positive association between maternal sensitivity and language development that was unique to boys. This important finding is indicative of a consistent cross-cultural effect whereby boys are more susceptible to changes in their environment during infancy. It is also possible that, in a higher risk sample with higher levels of clinically significant depression, the hypothesised interaction between depression and sensitivity on language outcomes may have been observed in boys.

3.5.6 Do sociocultural factors in India create a different dynamic between maternal depression, caregiving and infant cognitive development in India?

In a recent Lancet review, it was suggested that each of the accumulated adversities often observed in LMIC settings would be more detrimental to child development than each of those adversities occurring in isolation (Black et al., 2017). Stein et al., (2014) also reported consistent small to medium associations between postnatal depression and cognitive development in HIC and LMIC settings, particularly when symptoms are persistent. In view of this, it was expected at the outset that the effect of early postnatal and chronic

depression would be stronger in India than has been observed in HICs where risks are less likely to co-occur (Herba et al., 2016). However, it is also plausible that the interaction between individual socio-economic risk factors and maternal depression functions differently in LMIC settings (Garman et al., 2019), particularly as regards the role of maternal caregiving.

3.5.5.1 The key role of maternal caregiving in HIC literature.

The expectation that maternal depression will have a negative effect on development is largely predicated on the notion that an infant's immediate environment makes a significant contribution to that development. According to Sameroff's (2010) unified theory, early development relies largely on the provision of *other-regulation* provided by an external source, a role generally presumed to be fulfilled almost exclusively by the mother. *Other-regulation* has elsewhere been conceptualised as *nurturing care,* defined as stable home environment that provides a sense of emotional support and responsiveness, appropriate developmental stimulation, opportunities for play and exploration, and adequate protection from adversities (Britto et al., 2017).

There has been a substantial debate surrounding the applicability of the concept of maternal sensitivity in non-western cultures. However, Ainsworth's original definition and scale (Ainsworth, Bell, & Stayton, 1974), which was developed following observations of a sample in Uganda, appears to be suitable for the observation of sensitivity across cultures as it facilitates the rating of *"culture-specific behavioural manifestations that serve the universal function of making sure that infants receive what they need to survive"* (Mesman et al., 2017, p.10). Mesman et al. (2017) note that issues have arisen primarily from more recent measures of sensitivity which have utilised criteria that apply almost exclusively to western contexts, in particular, the addition of components regarding maternal warmth and affectivity (Cheah, 2016; Cheung & Elliot, 2016). These criteria have created an inflexible focus on the *form* of sensitivity, rather than its universal *function.* With this in mind, the current study focused only on the sensitivity to non-distress scale from the NICHD coding system, rather than creating composite index from a set of codes which typically include a rating of positive regard (warmth) shown within the interaction (NICHD ECRN, 1999). Additional steps to ensure cultural validity included training local experts to rate interactions with an emphasis on rating each scale based solely on the outcome of a mother's behaviour, rather than the specific form her behaviour took. The present finding that sensitivity to non-distress predicted language development in boys is consistent with

HIC literature and provides further support that, when coded appropriately, maternal sensitivity is a universal concept that bears considerable cross-cultural relevance.

In the current study, a significant and negative bivariate correlation was found between early postnatal depression and maternal sensitivity in the full sample, although the strength of the correlation was very weak (r = -.13) and there was no evidence of a moderation effect in the context of depression on development. Thus, it would appear in the current sample that maternal depression did not compromise the provision of nurturing to the extent that cognitive development was also impacted. There are a number of possible explanations for this. The first is that, as already discussed, the mild-moderate depression captured by the self-report screening questionnaire was not severe enough to represent a significant threat. However, it is also possible that certain socio-cultural factors unique to India, and other similar contexts, played an important role either in buffering infants against the impact of, or representing additional risk independent to, maternal depression. It is also possible that these factors may be specific to urban populations in India, as suggested by similar null findings reported regarding the effect of antenatal depression in an urban population by Thomas et al. (2020), and the contrast with findings from a more rural sample (Patel et al., 2003).

3.5.5.2 The potential role of economic deprivation and general family stress in India

Firstly, it is possible that the increased frequency and co-occurrence of adversities in LMICs may not simply exacerbate the effects of one another as previously thought. Instead, the presence of increased adversities may be themselves affecting caregiving and development such that the impact of postnatal depression is negligible. Controlling for deprivation indicators within the current sample does not account for cross-cultural disparities between HICs and LMICs and several studies have found that differences in maternal sensitivity between ethnicities were most strongly predicted by differences in SES (Heng et al., 2018; Mesman, van Ijzendoorn, & Bakermans-Kranenburg, 2012). These findings complement the family stress model (Conger & Donellan, 2007) which posits that socioeconomic disadvantage provokes an increase in familial stress which in turn disrupts sensitive caregiving. Additional findings have shown that caregiving behaviours typically found to promote cognitive development can be strongly affected by competing pressures and demands which divert maternal resources away from the child, and which are typically higher in the context of economic deprivation (Cecil, Pickles, Hill, & Sharp, 2017; Crockenberg & Leerkes, 2003b). So, while there is evidence that maternal depression significantly affects caregiving behaviours that promote cognitive development in HIC

settings, it is possible that it accounts for little unique variation in more adverse contexts. The finding that SES was a much stronger predictor of cognitive and, to a lesser extent, language development in boys than maternal depression in the current study supports this hypothesis. It is possible that boys, being more reliant on external scaffolding for development (Azak, 2012), may be more vulnerable to the impact that lower SES has on the broader caregiving environment.

3.5.5.3 The potential role of culture and family structure in India

Alternatively, it is possible certain socio-cultural characteristics may make families in India more resilient, both to the impact of maternal depression on caregiving behaviour, and to the impact of impaired caregiving on children. Chadda and Deb (2013) define India as a *collectivist* society which prioritises group goals and promotes family cohesion, interdependence, cooperation, and solidarity over more *individualistic* goals such as self-reliance and independence. As a result, the authors state that families in India are far more involved in caring for their members. The difference between India and many HICs is perhaps best seen in the traditional joint-family model found in India compared to the more common nuclear-family model found in many HICs. Against this backdrop it is possible that the nurturing care provided within families and the emphasis on shared responsibility may buffer the child against the effects of maternal depression on maternal sensitivity and other individual challenges. In this context then, it may wrong to place as much emphasis on the role of the mother in promoting development. Not because she does not play an important role, but because a higher degree of responsibility is shared between other family members and the support offered by families and communities is more likely to offset the effects of factors such as maternal depression (Heng et al., 2018).

3.5.7 What drives sex-differences in susceptibility to maternal sensitivity?

The current study supports the existing literature to an extent by providing evidence that boys' language development may be more reliant on sensitive caregiving than girls', although this finding was not evident for cognitive development. This finding also represents an extension of the literature in that this is one of the first studies to replicate this pattern of results in a LMIC.

An important question is why the impact maternal sensitivity was more evident in the language subscale than the cognitive subscale. It is notable that the mean language subscale score (M = 98.34, SD = 11.72) was both higher and demonstrated more variance than the cognitive subscale score (M = 93.90, SD = 8.29). Clinician observations indicate

that infants in India are typically less likely to be given toys to play with, and it was also observed by research assistants that, when asked to bring a familiar toy as part of the mother-infant interaction task, a high proportion of mothers brought some kind of household implement rather than a toy per se. Thus, it is possible that the reduced mean cognitive subscale score may have be due, at least in part, to a lack of familiarity in how to manipulate the objects used in the task rather than a genuine deficit in innate cognitive ability. In contrast, the language items test an area of development that is likely to be heavily impacted by the level of interaction and speech directed toward the infant. This may explain why the impact of maternal sensitivity was more evident in the language subscale scores, and why the expected impact on the cognitive subscale, based on HIC findings, was not observed.

Sex differences in susceptibility to the effects of maternal sensitivity are thought to be driven by early differences in temperament, specifically the development of affective and attentional self-regulation which is key for cognitive development (Bridgett et al., 2009). Early on this capacity is largely facilitated in all infants by the scaffolding provided by caregivers (Sameroff, 2010), but there is evidence that boys' ability to regulate their attention and emotion is more reliant on relational factors and the scaffolding provided by a sensitive caregiver (Donald et al., 2019; Grace, Evindar, & Steward, 2003). A meta-analysis of gender differences in temperament found a large and significant difference between boys' and girls' effortful control, including specific dimensions such as attention span, attention change and inhibitory control. The authors concluded that girls appear to show a significantly stronger ability to regulate and allocate their attention and inhibit impulses than boys (Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006). In one specific example, male infants were found to have greater difficulty maintaining affective regulation and spent more time looking to their mother, while female infants were more likely to spend time exploring and were able to self-regulate by focusing on objects (Weinberg, Tronick, Cohn, & Olson, 1999). Where male infants are unable to adequately self-regulate and mothers are likewise unable to provide sufficient external regulation, they are likely to miss out on key developmental opportunities.

Substantial evidence for increased male susceptibility to the effects of caregiving is provided by studies which have found sex-specific effects of maternal depression and impaired caregiving on cognitive development (Azak, 2012; Hay et al., 2001; Milgrom et al., 2004; Murray et al., 1993, 1992, 2010; Sharp et al., 1996). However, further evidence is apparent in the literature exploring the effects of non-parental childcare on development.

Evidence suggests that that boys tend to be more adversely affected by lower quality day-care than girls in several domains, including cognitive and academic achievement (Brooks-Gunn, Han and Waldfogel, 2002; Crockenberg, 2003). Boys who experience 30 hours per week of non-parental childcare are more likely to be insecurely attached than girls at 15 months (NICHD ECRN, 1997), and even in high quality care settings, increased levels of cortisol have been observed in both boys and girls, but only in boys does this predict poorer internalising and externalising outcomes (Tout, de Haan, Campbell, & Gunnar, 1998). These findings illustrate gender differences in regulatory ability and a heavier reliance by male infants on maternal care that provides a sensitive and effective form of external regulation (Vortruba-Dryzal, Coley, & Chase-Lansdale, 2004).

3.5.8 Further findings regarding the role of specific family-based variables.

The finding that the presence of an alternate caregiver (ACG) predicted poorer outcomes in boys, but not girls, may also be better understood in this context. While ACGs are typically family members, the non-parental element of ACG care means that these findings still bear some relevance. It is possible that ACGs, like day-care providers, are unable to provide the same quality of scaffolding and regulation as mothers, leading to poorer developmental outcomes for male infants. This may appear to stand in contradiction to the earlier discussion regarding the potential benefits of the joint-family context for offsetting the impact of maternal depression on caregiving. However, it is possible that while increased family involvement may act as a general buffer against the impact of individual challenges, there may also be a unique contribution of certain elements of shared caregiving that negatively impact boys. The potential interplay of gender and family-level variables is further shown by the consistent finding in this study that multiparity predicted significantly worse cognitive and language outcomes in girls but not boys. Donald et al. (2019) reported a similar gender-based effect of primiparity on language outcomes in South Africa. Being lower down the biological birth order is thought to inhibit development because parental resources become diluted as family size grows (Heiland, 2009). Although it is unclear why this effect is restricted to female infants in the current study, studies have now shown birth-order effects apply to both biological and social ranks within the family (Kristensen & Bjerkedal, 2007). Given the established presence of gender-bias in favour of male infants in India (Barcellos, Carvalho, & Lleras-Muney, 2014) it is possible that the social rank given to a male infant lower-down the biological birth-order may reduce the impact of resource dilution. As this was not the focus of the current study, it is not possible to answer these

questions with any certainty, and so future studies should be developed to tease apart the role of gender, family and shared caregiving in this context.

3.5.9 Strengths and Limitations.

The study was characterised by a number of design strengths. This was a prospective study of a large consecutive sample recruited from antenatal clinics serving a defined urban area in Bangalore, India. This allowed for the investigation of the prospective relationship between maternal depression and infant cognitive development. Although characterised by relatively low socioeconomic status, the community sample is generally representative of the population within the state of Karnataka. The version of the EPDS used in this study subject to a rigorous, committee-based translation process and validated in the local population against a psychiatric interview, with different cut-off values established for the immediate perinatal period and the later postnatal period. This approach allowed for different response styles that were apparent at these different time-points. The prospective design meant that EPDS data was available at three time points during pregnancy and four time points post-birth, allowing for adequate control of antenatal depression and the creation of a relatively robust chronic depression factor across the first 2 years of infants' lives. Additionally, cognitive outcomes and mother-child interactions were independently observed and assessed by researchers using well established and widely used measures, avoiding the potential for error arising from shared method variance with the self-reported depression scale. Finally, controlling for the effects of a large array of covariates, including antenatal depression, provides a high degree of confidence in the current findings.

Unfortunately, there was no measure of nutrition in the study and so it was not possible to control for the potentially confounding effects on maternal depression and cognitive outcomes. Another key limitation is the use of the EPDS, a screening instrument that does not provide a clinical diagnosis of depression. Nevertheless, while a tool providing clinical diagnosis may have led to more precision in depression measurement, the current approach allows for comparisons of the effects of mild-moderate depression in the current sample with more severe depression in other samples. Additionally, due to a delay in funding, a number of infants were not able to complete the 6-month assessment, meaning that some mother-infant interactions were assessed at 6 months but others at 12 months. It would also have been beneficial to have completed more observations of mother-infant interactions to investigate whether some developmental periods provide more of a challenge to parenting than others. Given the timing of the caregiving observation (6-12 months) and developmental assessment (24 months) it would have been possible for the

sensitivity exhibited by the mother to have changed significantly in the intervening period and for any impact to have been attenuated.

A unique challenge in sample retention in this context is that it is common practice for the women in the cohort to migrate to their maternal home towards the end of pregnancy and to remain there for the first 6-9 months postpartum. Whenever possible, research assistants travelled to these homes to complete assessments, but in many instances the participants had moved too far from the research site and so could not be followed up at that time. Additionally, many participants also travel to their native homes for various festivals which, despite the best efforts of the research team to reschedule appointments, meant that a large number of children were unavailable at the appropriate age and therefore missed assessments. This contributed to the lower sample size for the early postnatal depression analyses. Although follow-up and attrition rates were at a similar level to similar studies in this context, follow-up at 6 months was particularly low. This was because a delay in the release of funding meant that assessments could not be started on time, and many infants had passed the relevant developmental stage by the time assessments began. This may have had impacted the chronic depression variable as it included EPDS data at this time-point. However, the use of maximum likelihood estimation provides a robust method of imputation to account for missing data and so this is not expected to have significantly affected the final variable.

A final limitation is that although the version of the EPDS used was validated in the current population there is very little research that has explored how the EPDS functions in India, or any LMIC, relative to how it functions in HICs like the UK. The validation against diagnostic criteria provides broad reassurance that it is functioning similarly at a construct level, but it does not reveal anything about the more nuanced elements of cross-cultural functioning. This will be explored in the following sections of this thesis and the present results will be discussed in the light of those findings.

3.5.10 Implications.

The key finding from this study is that mild-moderate self-reported early postnatal and chronic depression is either not or only very weakly associated with impaired cognitive or language development in infants at 2 years in a low-income urban community sample in South India after adjusting for covariates. This has positive implications for child development in resource constrained settings. Maternal postnatal depression is a relatively common disorder and so the indication that it is not impacting on later infant cognitive

development should come as a relief to many. However, a degree of caution is required in interpreting this finding. Firstly, this is only the second study to investigate this relationship in India and the other published study (Patel et al., 2003) did find an association between early postnatal depression and cognitive development at 6 months. While the time of the developmental assessment represents a significant difference between the two studies, more studies are required to establish whether the pattern of results presented in this paper and many others across HIC and LMIC settings is replicated consistently within India.

Secondly, although maternal depression assessed using self-reported screening instruments may not predict poor child outcomes, there have been consistent findings in HIC studies and a few studies in different LMIC settings that have provide evidence of a significant effect of more severe, clinically diagnosed depression. This finding should not, therefore, be interpreted as meaning that postnatal depression does not represent a risk for child development. Instead, it is important for future studies to investigate the impact of clinically diagnosed postnatal depression on child outcomes in a community setting. For clinicians, this finding implies that while screening tools may be useful for identifying at-risk mothers, further assessment to determine the presence of clinical diagnosis will be useful in delineating those mothers and infants who are most at risk. This an important implication in LMIC settings as it means that the limited resource available can be more strategically allocated.

Maternal sensitivity has been highlighted as a key pathway through which depression affects development but there is no evidence of this in the current sample. There are a number of reasons why this may the case. Firstly, it may be because the depression experienced by mothers was not severe enough to impair caregiving. Secondly, it is possible that the broad collectivistic principles followed in India and the traditional joint-family model may mean that infant development is less mother-centric, and that certain sociocultural factors unique to India may buffer infants and mothers from the potential impact of maternal depression. However, maternal sensitivity was found to promote language development in male infants, indicating that maternal caregiving quality is still important. This is consistent with established HIC research and implies that boys are more reliant on the regulatory scaffolding provided by a sensitive and responsive caregiver. This leads to a third possibility, that factors related to socioeconomic disadvantage represent a greater risk to caregiving and child development than maternal depression. It is therefore important for future studies to look to disentangle the roles of sociocultural factors and factors associated with lower socioeconomic status in order to better understand the

dynamic of the relationship between maternal mental health, caregiving quality, and infant development.

3.5.11 Conclusion.

This is only the second study to explore the prospective relationship between early postnatal depression and infant cognitive development in India, and the first to explore the impact of chronic depression and the role of maternal sensitivity. Adjusted estimates showed no association between early postnatal or chronic exposure to maternal depression and impaired cognitive or language outcomes. There was also no evidence that maternal sensitivity mediated or moderated this association, but maternal sensitivity did significantly predict language development in boys. The findings indicate that boys are more reliant on the external regulation provided by mothers and that factors related to economic deprivation for boys and parity for girls may be more relevant to developmental outcomes than self-reported maternal depression in this setting.

Chapter 4: Cross-cultural Measurement Issues in Mental Health Research

4.1 Introduction

Historically, there has been a widely held assumption that a robust approach to measure development in a western context negates the need for any further work or thought when that measure is used in a cross-cultural setting. According to Rogler (1999) this position effectively labels the local culture an irrelevant and extraneous intrusion to the research at hand. While the methodology used in cross-cultural research has developed considerably in recent years, concerns remain regarding what appears to be a relatively persistent belief that the robust techniques used to establish the validity of a given measure in one context are enough to ensure that it is valid in another (Morris, 2018). Inevitably, this threatens the ultimate goal of cross-cultural research in mental health, which is the valid comparison of psychological phenomena around the world (Canino, Lewis-Fernandez, & Bravo, 1997). In this light, it is imperative that researchers demonstrate desirable psychometric properties regarding the cross-cultural comparability of target measures before any comparisons are made (He et al., 2017). Researchers have a responsibility to clearly articulate and, where possible, minimise the bias that may contribute to non-equivalence in their studies (Morris, 2018).

4.1.1 Use of the Edinburgh Postnatal Depression Scale in India

The Edinburgh Postnatal Depression Scale (EPDS; Cox, Holden & Sagovsky, 1987) is the most widely used screening questionnaire for postnatal depression in global research (Boyd, He, & Somberg, 2005) and has been translated into over 60 languages (Cox, 2019). As such, it is a measure which warrants close examination in terms of its validity across different cultures. The scale was developed and validated in the United Kingdom and optimal sensitivity and specificity was reported at a cut-off of 13 or higher for the detection of probable depression, and 10 or higher for the detection of possible depression. Although many subsequent validation studies have been carried out, a considerable number of articles sampling a remarkably diverse set of countries and cultures report the use of the original cut-off with no clear reasoning or justification. For example, in a recent meta-analysis of postnatal depression in India, 29 out of the 38 studies used the EPDS (Upadhyay et al., 2017), but only 8 validated the cut-off used or cited another validation study within India. Additionally, in the current study, validation of the EPDS against a clinical diagnostic interview in the antenatal and postnatal periods found markedly different cut-points at different stages in the perinatal period, with optimal sensitivity and specificity found at a

threshold of 3 or above for the antenatal and immediate postpartum period (8 weeks), and a threshold of 10 or above for the late postnatal period (6-24 months) (see Appendix 4). This shows that even within India, there are significant differences in how the EPDS is functioning at different time-points and suggests, at the very minimum, that the EPDS should only be used in a context where it has been previously validated. This is emphasised by two systematic reviews of studies which have used the EPDS alongside clinical diagnostic measures in different cultural settings (Gibson, McKenzie-McHarg, Shakespeare, Price, & Gray, 2009; Srestha. Pradhan, Tran, Gualano, & Fisher, 2016). The authors reported that there was substantial heterogeneity in the sensitivity and specificity of the standard cut-off points and that this difference was particularly marked between English-language (sensitivity = 76-100%, specificity = 70-99%) and non-English-language versions (sensitivity = 49-100%, specificity = 34-100%). These findings suggest that the EPDS may not be functioning equivalently in every cultural and developmental context and that a more culturally sensitive approach to using and interpreting this scale is needed.

4.1.2 A culturally sensitive approach to cross-cultural research

Sensitive cross-cultural measurement requires an approach that emphasises the minimisation of bias and the evaluation of invariance. Bias can be defined as the nuisance factors that jeopardise the validity of instruments applied in different cultures, while invariance can be defined as the opposite of bias and refers to the level of comparability between cultures (He & Van de Vijver, 2012). More specifically, bias refers to systematic and replicable measurement error and is said to be present when differences between groups in a given construct (e.g., depression) captured by a particular scale (e.g., EPDS) do not correspond to differences in the underlying trait (Van de Vijver & Tanzer, 2004). An important distinction at this point is that both bias and equivalence do not refer to intrinsic properties of a particular measurement tool, but to characteristics of a particular cross-cultural comparison of that instrument. For example, if the EPDS reveals bias in a comparison of UK and Indian respondents, it may not show the same bias in a comparison between Indian and Chinese respondents. However, for the sake of comparison it can be beneficial to consider bias in relation to a pre-defined anchor point. Thus, in continuation of the prior example, as the EPDS was initially developed and validated in the UK, scale use in this population could be considered as the “gold-standard” and any cultural discrepancies considered in terms of what is driving differences in relevant comparator groups.

4.2 Taxonomy of bias

Bias is typically organised into a three-tiered taxonomy including construct, method, and item bias (Byrne & Watkins, 2003; He & Van de Vijver, 2012; Van de Vijver & Poortinga, 1997; Van de Vijver & Tanzer, 2004).

4.2.1 Construct bias.

Construct bias indicates that the construct being measured is not identical across cultural groups. It can occur where there is only partial overlap in the definition of a construct, or where not all behaviours relevant to a given construct are properly assessed by the instrument. A key example of this is in the assessment of depression between western and non-western settings. Parker, Cheah and Roy (2001) found that Chinese depressed outpatients highlighted somatic symptoms as key complaints, while Australian comparators were more likely to speak in terms of depressed mood and cognitive anxiety symptoms. Elsewhere, it has been reported that somatic symptoms may predominate self-reports of depression in many non-western settings, including India (Varma, Chandra, Thomas & Carey, 2007). Thus, a questionnaire such as the EPDS, which is designed explicitly to exclude the potentially confounding somatic complaints that commonly occur in the perinatal period (Cox, Holden, & Sagovsky, 1987), may not provide a holistic assessment of depression in India (Srestha et al., 2016). Where this is the case and is unavoidable, Van de Vijver and Tanzer (2004) suggest continuing with the comparison while acknowledging the incompleteness of the measure for the non-western group.

4.2.2 Method bias.

Method bias refers to the nuisance factors that arise from sampling, instrument, and administration bias. Sampling bias refers to the incomparability of samples due to cross-cultural variation in sample characteristics that have a bearing on target measures. In the current study, it is expected that differences in specific sample characteristics, and more generally the combination of poorer provision of perinatal mental health services at a national level, precipitating a lower awareness of its importance at a personal level (Fisher, de Mello, Izutsu, & Tran, 2011), and the increased exposure to socio-economic adversity in LMIC settings (Black et al., 2017), will partially drive differences in the level of depression reported between groups. Instrument bias occurs where scores are influenced by different levels of familiarity with stimulus materials, response modes or response procedures. A common example of this rests in the Likert-type scaling format, which is generally more familiar to groups drawn from western settings. Instrument bias can also be understood in

terms of differences in response style, with people of low socioeconomic backgrounds from collectivist cultures being more likely to agree than disagree with a proposition (Harzing, 2006). Finally, bias can arise as the result administration problems. Ambiguous instructions and communication problems can have a significant effect on responses given by participants, as well as the mode of delivery. For example, bias may arise in comparisons between HICs and LMICs because, due to differences in literacy levels, a questionnaire may be self-administered in one group and researcher administered in another. In this example, bias may operate at two levels in participants who receive the researcher administered scale. First, these participants may feel pressured to give a response quickly and therefore not take as much time to carefully consider their answer (Molina, 2017). Second, participant responses may be unduly influenced by a social desirability effect, leading to potential under-reporting of symptoms viewed to be negative (He & Van de Vijver, 2012).

4.2.3 Item bias.

Item bias, also known as differential item functioning (DIF), occurs where individual items have different characteristics across cultures. This disparity may originate where the diversity of sociocultural contexts to which each group is exposed is not properly accounted for during the development and/or translation of the scale. It can arise from poor translation, inapplicability of item content in different cultures, words with ambiguous connotations, or from items that trigger additional traits. Items are considered to be biased by DIF when, *"two individuals who are equal on the underlying quantity of interest nonetheless have unequal probabilities of providing the same answer"* (Hopkins & King, 2010, p.2).

4.2.4 Response set bias.

Going one step further, DIF can also produce bias in how different groups understand and use item response sets (e.g., "Not at all, Sometimes, Very Often"). Although similar to item bias, response set bias represents a unique challenge whereby a specific response option to a given item does not represent the same level of the symptom described in that item across groups. This is particularly important because it is not as easily remediated as item bias through careful and sensitive planning. Furthermore, because they tend to be relatively cheap and quick to administer in large populations, self-report questionnaires predominate in cross-cultural mental health research.

Self-report questionnaires, such as the EPDS, often utilise ordered response categories which participants use to rate their level of health. These self-assessments depend upon

the objective reality of an individuals' true (latent) health and their subjective view of what it means to be above or below given response thresholds (Van Soest, Delaney, Harmon, Kapteyn, & Smith, 2011). Problems arise when individual heterogeneity, resulting from the context or culture that different individuals live in and experience, cause those individuals to interpret and utilise these categories or thresholds in different ways (Paccagnella, 2013). When this behaviour is systematic across groups or cultures it can produce biased or distorted results (Crane, Rissel, Greaves, & Gebel, 2016). Figure 4.2.4 illustrates how differences in the placement of response thresholds may lead to misleading results.

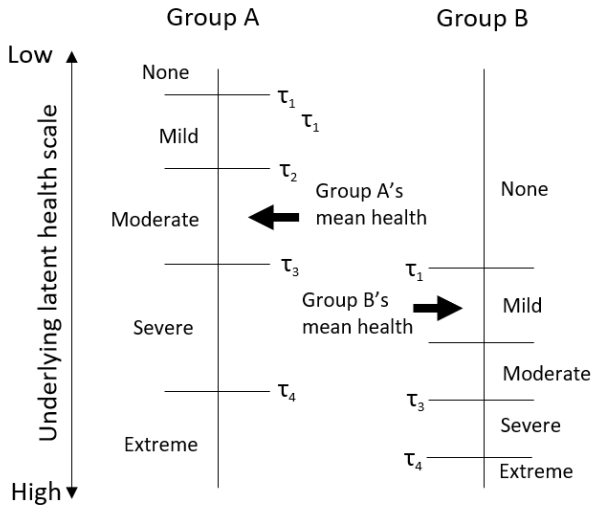


Figure 4.2.4: Illustrated example of DIF. Individuals in Groups A and B are asked to self-report their own level of health problems. How the average individual in each group divides the underlying latent into the five response categories is represented by t1 to t4. DIF is portrayed by variation in the placement of the thresholds across the two groups. Based on self-ratings, group A would be assumed to have more health problems, even though their actual level is lower (Figure adapted from King et al., 2004).

An example of this can be seen in the results of a comparative health study in India (Sen, 2002). Self-reports indicated that residents of Kerala were less healthy than those in the rest of India, despite objective measures indicating that the opposite is true. This may be because residents of Kerala have significantly less exposure to mortality and morbidity than residents of other states, such as Bihar. Individuals with higher standards of health are likely to consider the same health issue as more serious than someone with lower standards of health. They may, therefore, rate themselves as less healthy than someone of the same objective health state but who has lower standards of health. In this case, the quality of healthcare in Kerala has caused a shift in the healthcare standards and the thresholds that residents use to rate their own health, thereby providing results that indicate a lower relative standard of health than is objectively true.

4.3 Taxonomy of invariance

Similar to bias, measurement invariance is typically defined in a three-tiered taxonomy (Chen, 2008; He & Van de Vijver, 2012; Putnick & Bornstein, 2017). An important distinction is that the taxonomy of invariance is linked in a hierarchy whereby obtaining each level is dependent on obtaining the one before (Van de Vijver & Tanzer, 2004). Although there are several other levels of invariance that can be tested, they are not required for comparing scores across groups and are only theoretically meaningful in certain contexts (Milfont & Fischer, 2010). It should also be noted that the term "measurement invariance" is used interchangeably with "measurement equivalence" in the literature (Morris, 2018).

4.3.1 Construct invariance.

The first level of measurement invariance is construct, or configural, invariance. Construct invariance means that the same construct is measured across all groups studied and implies the universal validity of the underlying psychological construct in those groups. Without this, there is no basis for comparison. Non-equivalence at this level typically occurs when construct bias is present, as described in section 4.2.1.

4.3.2 Metric invariance.

The second level of measurement invariance is measurement unit, or metric, invariance. Metric invariance means that each item contributes to the latent construct to a similar degree across groups. When obtained the scale has the same measurement unit in each group. If metric invariance is obtained within-group differences or effects can be compared between groups. Speaking in terms of the current study, if measurement unit invariance is shown for the EPDS between the UK and India, the relationship between postnatal depression and age within the UK can be compared with the same relationship within India.

4.3.3 Scalar Invariance.

The highest level of invariance is scalar invariance. This level is obtained when a scale has the same unit and the same origin (or intercept) across groups. A clear example of difference between metric and scalar invariance is seen in the measurement of temperature. Kelvin and Celsius both have the same measurement unit and so exhibit metric invariance, but the origin (intercept) of Kelvin is shifted 273 degrees beneath Celsius, and so they do not exhibit scalar invariance. In contrast, two temperatures both measured in Celsius do exhibit scalar invariance.

At the level of scalar invariance, measurement is assumed to be completely bias free; mean differences in the latent construct capture all mean differences in the shared variance of

the outcomes. Another example of non-invariance of an item intercept would be when mothers in one culture cry more, but that because the scale intercept is shifted lower in that culture, increased crying is not related to increased levels of postnatal depression relative to the other group.. In the absence of scalar invariance, direct cross-cultural comparisons are only possible where the difference in intercepts of the scale between the two groups is known. For example, the difference in origin between Kelvin and Celsius is known to be 273 degrees and so the values from the two scales can be converted so that they are equivalent. In reality, the actual difference between the origin of scale in one group compared to its origin in another group is rarely known. This means that additional information is required to determine that difference and then convert the measurement to the same origin in both groups (Van Soest et al., 2011).

4.4 Approaches to evaluating invariance and minimising bias.

There is a wide array of techniques available to minimise bias that can be utilised in the planning and development phase of a research project. Where an entirely new scale is being developed, decentering, which involves simultaneously developing the same instrument in several cultures, spares researchers from the rigidity of trying to force one version to match the potential idiosyncrasies of the other, instead allowing each one to be modified to find the best balance (Erkut, 2010). This minimises construct bias by reducing the potential for a scale to be more relevant to one specific culture. Unfortunately, when using an established scale like the EPDS, this is often not possible, as changing the original, validated version is not an option. In such cases, more emphasis is placed on minimising bias in the approach taken during translation, administration, and analysis.

Clearly, how a scale is translated will have a significant effect on the equivalence of that measure to the original. Very often, simple linguistic or semantic equivalence will not yield invariance and a more nuanced approach that seeks to establish functional equivalence is required (Pan & Fond, 2014). Construct bias may be further minimised if the translation is carried out as part of an iterative forward- and back-translation process conducted by a committee of experts in the field of study, as well as the local culture and language (Van de Vijver & Tanzer, 2004). Once the scale has been translated, He and Van de Vijver (2012) recommend a number of strategies to reduce method bias, including extensive training of administrators and comprehensive standard protocols for administration, flexibility in how the scale is administered in different settings, clear instructions for participants, and the collection of subject and context data to check for the effects of potential confounding

variables. These techniques are further supported in the specific context of the EPDS by Srestha et al. (2016) who note that accurate detection of perinatal mental health disorders is predicated on a scale's comprehensibility to the target population, and that this requires a culturally sensitive approach to translation and administration.

However, even when the guidance on minimising bias is followed religiously, it should not be assumed that a scale has achieved measurement invariance between two groups. Moving beyond this assumption is important for providing confidence in cross-cultural research. The following sections will outline two distinct approaches that can be used in concert to deliver a culturally sensitive approach to measurement in a cross-cultural context. The first approach, concerned with evaluating measurement invariance, involves the use of Confirmatory Item Factor Analysis to test for each level of the taxonomy of invariance described earlier. The second, the anchoring vignette approach, is a relatively novel methodology that allows for the detection and elimination of item response bias post-hoc.

4.5 Measurement invariance testing

Measurement invariance testing assesses the psychometric properties of a scale across groups or across time. It is a well-established procedure that can be tested using an item-response theory (IRT) framework or a structural equation modelling (SEM) framework. The current project focuses exclusively on the SEM framework as this is the dominant approach in the literature (Putnick & Bornstein, 2016). The SEM approach uses multi-group or longitudinal confirmatory factor analysis (CFA) to successively test model fit at each level of invariance as outlined in the taxonomy above (configural, metric, scalar), and worsened fit is taken as evidence of non-equivalence. Additional constraints are added to the model at each level in a test of increasingly strict invariance (Chen, 2008, Milfont & Fischer, 2010). In the first step, configural invariance is tested by constraining the factorial structure to be the same across groups, and invariance is achieved if the same item is significantly associated with the same latent factor in each group. In the second step, metric invariance is tested by constraining factor loadings to be equal between groups. If the model fit is not significantly worse, this indicates that the items are contributing to the factors, or underlying traits, in the same way in each group. In the third step, scalar invariance is tested by constraining item thresholds to be equal between groups. If the model fit is not significantly worse, this indicates that the item responses relate to the factors, or latent traits, in the same way across groups. Where both metric and scalar invariance is achieved, scores from different

groups are assumed to have the same measurement unit and origin and latent means can be compared across groups. Where scalar invariance is not achieved, further information regarding the origin of the scale would enable the measurement to be converted to the same origin. In this situation, the anchoring vignette methodology can be used to compliment the measurement invariance findings and minimise the bias identified.

4.5.1 Partial Measurement Invariance.

Full measurement invariance is an extremely strict goal and is unlikely to hold in practice. In response to this, partial measurement invariance was introduced by Byrne, Muthen, and Shavelson (1989). In this case, a subset of parameters at each level are constrained to be invariant, while another subset is allowed to vary across groups, thus allowing appropriate cross-cultural comparisons where full measurement invariance is not achieved. Vandenberg and Lance (2000) clarified three criteria for using partial invariance. First, full configural invariance and partial metric invariance should be established before testing any further partial invariance model (e.g., partial scalar). Second, partial metric invariance is only permitted if parameters relaxed to vary across groups apply to a minority of indicators, or items. Finally, there should be a strong theoretical rationale for allowing parameters to vary between indicators.

4.6 Anchoring Vignettes

4.6.1 Introduction to anchoring vignettes

Scalar non-invariance, discussed above, is a significant threat to the validity of cross-cultural comparisons and is often driven by DIF in item response sets. Due to the typically abstract and non-concrete nature of the terms used in response items, this bias is not readily remedied by even the most careful translation. Therefore, a more novel approach is required. The anchoring vignette methodology was initially proposed by King, Murray, Salomon, and Tandon (2004) as an approach to enhancing the validity of cross-cultural research when measuring complex or abstract constructs. Although the method was primarily used in the field of political science, it has recently been utilised more widely, most notably in several large studies designed by the World Health Organisation (WHO), such as the World Health Survey and the Study on Global Ageing and Adult Health (Kowal et al., 2012).

4.6.2 Theoretical framework.

Anchoring vignettes (AVs) are very short narratives, normally one or two sentences long, that describe a hypothetical character and how they are feeling or behaving in relation to a

given construct (e.g., depression, quality of life, political efficacy). These narratives are translated into each target language so that the same vignettes can be shown to each group. In the methodology proposed by King et al. (2004), participants are typically first asked the self-assessment question(s) and then shown the vignettes and asked to rate the characters using the same scale (response set) which they used to rate themselves. After careful translation, the vignette characters are assumed to exhibit the same level, or severity, of the concept being measured in each group. As the true (latent) health of the vignette characters is considered to be "anchored" between groups, any systematic variation in ratings given between the groups can attributed to group differences in response style (Dasgupta, 2018; Salomon, Tandon, & Murray, 2004; Van Soest et al., 2011). By quantifying this variation, researchers are able to adjust group or individual responses to adhere to a common origin and scale across all participants, thereby making the data directly comparable (Gengler & Mitchell, 2018; Hopkins & King, 2010; Paccagnella, 2013). Molina (2016) concludes that this approach *"allows inferences about respondents' internal response scales that are otherwise completely unobservable to the researcher"* (p.299).

It is important to note that this method is designed to correct for response category DIF only. AVs will not fix issues where participants understand the stem question, or the vignette description, differently (King & Wand, 2007). King et al. (2004) suggest that it is reasonable to focus on DIF in the response categories as guidance on writing objective vignettes and self-assessment questions should be sufficient to remove DIF from those areas to a satisfactory degree. It is much more difficult to create DIF-free response categories as they are generally more abstract and refer to subjective feelings and attitudes that are very difficult to lay out in a concrete manner. The predication that vignettes will be free of DIF relies on the assumption of *vignette equivalence,* which will be explored in the next section.

4.6.3 Measurement assumptions

4.6.3.1 Response consistency

Response Consistency (RC) is the assumption that *"each individual uses the response categories for a particular survey question in the same way when providing a self-assessment as when assessing each of the hypothetical people in the vignettes*" (King et al., 2004, p.194). Similarly, Van Soest et al. (2011) describe it as the assumption that respondents use the same subjective thresholds in rating vignettes as they use when rating themselves. DIF may vary across respondents, and even within respondents across different survey questions, but it must be consistent across a single survey question and its

corresponding vignette(s) for a single respondent. This assumption is violated if respondents hold themselves to a different standard to vignette characters or use standards inconsistently across vignettes (Grol-Prokopcyck, Verdes-Tennant, McEniry, & Ispany, 2015). While discussions regarding why these violations may occur are few, Bago d'Uva, Lindeboom, O'Donnell, and Van Doorslaer (2011) suggest that certain external factors may influence self-assessment ratings but not vignette ratings. For example, non-working individuals may experience social or financial pressure to understate their own health but not that of hypothetical characters. In the context of the EPDS, a violation may occur if a participant perceives the coping ability of a hypothetical character to be greater than their own. If RC is violated, cut-points calculated from vignettes will not correctly adjust self-ratings (Grol-Prokopczyk et al., 2015).

4.6.3.2 Vignette equivalence

Vignette Equivalence (VE) is the assumption that "*the level of the variable represented in any one vignette is perceived by all respondents in the same way and on the same unidimensional scale, apart from random measurement error"* (King et al*.,* 2004, p.194). Put another way, VE is the assumption that the vignette evokes the same picture of the same underlying construct for all participants across all groups (Bago d'Uva et al., 2011). Any differences in how the vignette is perceived must be random and independent of the characteristic being measured. For the vignette to act as an anchor, the level of the variable must be perceived as the same by all respondents (Hopkins & King, 2010). If this assumption is violated, and the health of the vignette character is not held constant, there is no anchor about which to calculate DIF. Rather than observing how participants perceive and respond to the same states differently, we will be observing how participants respond to differently perceived states.

Violations of VE may occur if groups interpret the vignette texts in systematically different ways. For example, if a vignette character's annual medical visit is interpreted by residents of rich countries as a beneficial, preventive check-up, and hence indicative of good health, but is interpreted by residents of poor countries as a sign of frequent medical need and hence of poor health, then VE has been violated (Grol-Prokopczyk et al., 2015). Violations may also occur where vignettes that are used in multiple cultures are translated poorly. When cultural differences are not fully accounted for in the translation process, vignettes can end up portraying very different levels of health. The content of a vignette must carry the same weight in each of the cultures it is utilised in. Finally, if the construct being assessed is multidimensional, this assumption is likely to be violated if the vignette

formulation does not adequately capture the complete concept, thereby necessitating some imputation by the reader (Vonkova, Bendl, & Papakoanu, 2017).

4.6.4 Evidence for the utility of anchoring vignettes

4.6.4.1 Within-country comparisons

Although the AV methodology was originally designed for cross-cultural comparisons, the application has also been extended to include within-country comparisons.

Anchoring vignettes have been shown to detect DIF in a range of health-related domains, with emphasis placed on self-rated health (SRH) and quality of life (QoL) measurements. SRH can be assessed either as a single, general health question, or as part of a questionnaire that examines different domains of functioning. Grol-Prokopczyk, Freese, and Hauser (2011) detected DIF in responses to a single SRH question in the USA, with differences predicted by several demographic and health related variables, including gender, age, and education. Prior to correction, women's SRH was better than men's but following adjustment their SRH was the same or worse. Education was also found to drive DIF in SRH across Europe, with more highly educated individuals reporting more negative health outcomes. Here, DIF was masking inequality in healthcare access and utilisation, and the AV correction increased agreement between SRH and an objective indicator of health (Bago d'Uva, et al., 2011). Similar findings were reported regarding QoL ratings in Germany, where breast cancer survivors rated their own QoL and that of the AV characters as significantly higher, possibly because of changes to their internal frame of reference due to their own experience. Results of the analysis suggested that the apparent QoL reported by survivors and the general population is at least partially the product of DIF (Hinz, 2017).

Importantly, this phenomenon does not appear to be restricted to HICs. Despite objective indicators suggesting differently, SRH in South Africa is remarkably similar across different levels of wealth. Rossouw, Bago d'Uva, and Van Doorslaer (2018) found that wealthy individuals rated vignette descriptions as less healthy than poorer individuals, and after correcting for differences in response style, significant disparities in SRH emerged that favoured the rich. This same pattern is seen across 6 states in India, where individuals from more disadvantaged states consistently underestimated their health problems relative to individuals from more developed areas (Dasgupta, 2018). Similarly, Hanandita and Tampubolon (2016) reported that allowing for DIF based on level of education magnified the positive effects of education on health in Indonesia. Finally, Molina (2016) reported the extent to which within-country gender and education differences in six domains of SRH

were driven by DIF separately within four separate countries. Following correction, gender gaps were completely eradicated in China and England, and reduced in half of the measured domains for Indonesia and the U.S. The validity of the correction was supported by evidence from objective indicators that despite generally reporting worse SRH, women actually have much lower mortality rates than men. In contrast, the gaps in SRH by education level actually increased after correction, following the pattern seen in many AV studies and supported by the evidence from more objective indicators.

4.6.4.2 Cross-cultural comparisons.

Knott, Lorgelly, Black, and Hollingsworth (2017a) address both within-country and cross-cultural DIF in five domains of health-related QoL, including anxiety/depression. Within-country comparisons showed that education level drove differences in reporting behaviour in all five dimensions, while gender was a significant predictor of DIF in three out of five dimensions. Interestingly, while DIF adjustment resulted in a significant widening of the health gap between education groups, there was no significant change in the differences in QoL between males and females, which remained non-significant before and after adjustment. Cross-cultural analyses were carried out by investigating DIF between individuals born in Australia, other English-speaking countries, and Asian countries. After correcting for DIF, the difference between Australia and other English-speaking countries doubled in magnitude, and the difference between Australian-born participants and those born in Asia also increased substantially.

In an extension of her previous within-country research, Molina (2017) investigated differences in AV-adjusted SRH between Indonesia, China, England and the U.S. Consistent with basic objective indicators of health, the pattern in the raw data for U.S. and English participants to report worse SRH completely reversed after correction. Comparatively, Kang and Grol-Prokopczyk (2020) found that the trend for U.S. participants to report higher SRH than South Korean participants was reversed following AV correction. The indication that South Korean's had much higher standards of health and assigned much higher values to response thresholds suggests that the AV approach is sensitive to genuine differences in internal health standards and is not simply assigning higher standards to whichever western country is part of the comparison.

While the cross-cultural differences in reporting style above were detected in countries with starkly different sociocultural contexts, the AV approach also appears to be sensitive to DIF in cultures that are more outwardly similar. Kapteyn, Smith, and Van Soest (2007)

reported that substantial observed differences in work disability between the U.S and the Netherlands were due to systematic differences in response style, specifically finding that Dutch residents had a much lower threshold for reporting work disability. Mojtobai (2016) also found evidence of DIF between the U.S. and a number of European countries in single item self-ratings of depressed mood. U.S. participants rated themselves and the AV characters as more depressed than their European counterparts, but after correction, were less depressed than participants in seven out of nine European countries.

The utility of anchoring vignettes is further demonstrated in their adoption across a range of fields, including political science (King et al., 2004), education (Vonkova et al., 2017), personality (Mottus et al., 2012), and job satisfaction (Kristen & Johansson, 2008). The current review will not explore the use of AVs in these fields in detail but attention is drawn to a number of studies that highlight some of the strengths and limitations of the methodology. Vonkova, Zumarro, and Hitt (2018) tested cross-cultural comparability in students' reports of teacher classroom management using a very large sample (n = 310,000) across 68 country-regions. Results indicated significant variation in implicit standards across countries, and that correlations between ratings of classroom management and student outcomes changed significantly after correction. Most notably, correlations moved from negative to positive on a number of key policy variables (e.g. average test score), supporting more intuitive relationships. Weiss and Roberts (2018) showed that the use of AVs can demonstrably improve measure psychometric properties. Initial tests of the Big-Five Inventory of Personality showed that even basic configural invariance was lacking in a comparison between Rwanda and the Philippines, despite extensive attention to detail in the translation process. Following AV correction, scale reliability improved, as did the overall test information. The multi-group CFA for Rwanda and the Philippines also showed acceptable fit following correction, indicating that the AVs improved configural invariance.

These findings demonstrate the utility of, and the need for, a method such as anchoring vignettes that will account for the use of different response thresholds between groups. There is strong evidence that self-reported differences between groups in a number of health-related domains are subject to systematic biases in reporting behaviour and the AV correction brings responses more in line with objective indicators. DIF may be driven by within-country factors, such as education level, wealth and gender, or cross-cultural factors, such as differences in sociocultural context or economic development.

4.6.4.3 How many anchoring vignettes are needed?

Another important consideration in the utility of anchoring vignettes is the number of vignettes required for the method to be effective and the trade-off between the reduction of bias and increased survey costs. King et al. (2004) stated that their proposed parametric approach to analysis can function adequately on the basis of a single AV, although in practice they normally advise using more. More important than the number of AVs is the quality of the AVs and the discriminatory power they provide. Thus, AVs should be selected to divide up the distribution of self-assessment responses equally and effectively. Consequently, although only one vignette is required, multiple vignettes with a range of severities provide a higher measurement resolution. Additionally, the use of multiple vignettes allows for the inclusion of random effects in the modelling of threshold variation (King, 2009).

In the research discussed above, the vast majority of studies used 2-3 vignettes for each self-assessment item and followed the parametric approach to analysis set out by King et al. (2004). There were also a limited number of studies who used 5 AVs for each domain of self-rated health and, representing the more extreme ends of the continuum, one study which used a single vignette (Vonkova et al., 2017) and one study which used 10 vignettes (Dasgupta, 2018). In almost all studies, responses to both self-assessments and AV assessments were given on a matched 4-point or 5-point scale. The only exception to this was Hinz (2017), in which self-assessments and AVs were rated using a visual analogue scale from 0-100.

Interestingly, although the parametric approach allows for the inclusion of multiple self-assessment questions for the same underlying construct in a single factor analysis-type setup (King et al. 2004), only Vonkova et al. (2018) exploit this approach. Particularly in health-related research, studies appear to have almost exclusively focused on adjusting for DIF in single-item assessments. Although personality researchers have looked more often at multi-item scales, they have also typically utilised a non-parametric approach to rescaling self-assessment responses (He et al., 2017; Mottus et al., 2012; Marksteiner, Kuger, & Klieme, 2018). In view of this, the current study (see Chapters 5 & 6) appears to be one of the first to use the AV method with a multi-item scale specific to a single underlying construct, and it is almost certainly the only study to use a multi-item scale with distinct response sets for each item. While there is no specific guidance on what is required in this context, the logic of the method suggests that, if AVs detect DIF acting on response sets, then even if each item within a scale refers to the same construct, the effect of DIF on

distinct response sets is likely to be unique. In the example of the EPDS, where each item refers to an individual symptom of postnatal depression, this means that a separate set of AVs is required for each item that adequately describes the same symptom, can be sensibly answered using the same response scale, and effectively divides up the distribution of self-assessment responses.

4.6.4.4 Limitations to the Anchoring Vignette Approach.

Challenges and limitations to the AV methodology have also been noted. Hinz (2017) distinguishes between DIF driven by differences in sociocultural reference points and a judgement effect, whereby a particular group may be more likely to agree to an item (acquiescence) or use particular points of the response scale more often (e.g., extreme response style). In this case, it is possible that the judgement effect will override the reference point effect. However, AVs should theoretically detect differences arising from either. A more pertinent question may be how to account for individual, rather than group, differences in response style. Additional concerns are raised regarding the complexity that AVs add to a questionnaire and the suitability of the method for older populations (Poksinska & Cronemyr, 2017), as well as evidence that, when used longitudinally, AV ratings fluctuate non-directionally over time (Topp, Heesen, Augustin, Andrees, & Blome, 2020).

In a cross-cultural setting, Bago d'Uva et al. (2011) reported that while AVs did reveal a significant level of reporting heterogeneity of health within 3 developing nations, they did not correct results to the extent that objective indicators suggested. This may be because there is more noise in vignette data in developing countries, meaning that it will is more difficult to identify reporting behaviour from AV ratings. Vonkova & Hullegie (2011) also found that the utility of the AV method is sensitive to the choice of vignette, as specific vignettes had varying degrees of success in bringing the subjective and objective indices closer together. This indicates that measurement assumptions are not being met equivalently in all vignettes.

A final consideration is whether AVs improve measurable psychometric properties. He et al. (2017) evaluated the utility of AVs in enhancing the cross-cultural comparability of data across 16 countries, alongside a number of other techniques. While there was evidence that the AVs themselves were not bias-free, use of the AV ratings improved the internal consistency of personality subscales. AV correction also improved invariance to some degree but not to the extent that scalar invariance was achieved. They concluded that AVs

are not a *"magic bullet to achieve scalar invariance"* (p.13) but that, although rescaled scores based on AV ratings are not completely bias free, they are potentially less biased than raw scores. These findings were closely replicated by Marksteiner et al. (2018) who reported improved scale internal consistency and clearer factor structures following anchoring but also substantial ordering violations and only a marginal improvement to measurement invariance indices and never to the point where scalar invariance was achieved. It is noteworthy that 51 countries were included in comparisons, meaning that meeting the measurement assumptions will have been extremely demanding. It is possible that the adjustment was not as significant as expected because the vignette equivalence assumption was not met.

Critical thinking needs to be applied in interpreting and assessing the validity and meaning of AV results. While some studies have shown that the vignette-adjusted scores bring self-reports into a greater degree of agreement with objective measures, the majority of studies that utilise this method do not have an objective indicator available for formal comparison. Rather many studies rely on more general indicators of health from the literature and discuss the extent to which adjusted results line up with these findings. While this is certainly a valid approach, caution is needed, therefore, in the claims made regarding whether vignette adjusted scores are more valid than unadjusted scores.

4.6.5 Empirical support for assumptions.

Bago d'Uva et al. (2011) report that while some studies have tested the face validity of the AV methods by exploring whether they bring subjective and objective indices closer together, there has been very little, if any, formal testing of the underlying assumptions. The following sections will explore how the assumptions have been tested in the literature and the strengths and weaknesses of these approaches.

4.6.5.1 Response Consistency.

There are only a few examples in the literature of formal testing of the RC assumption. Van Soest et al. (2011) asked participants to give an objective frequency of how many drinks they would consume on a given occasion and then to rate their own drinking behaviour followed by rating the drinking behaviour of vignette characters. The distributions showed that participants tended to characterise vignette persons similarly to the way that they characterised their own drinking behaviour, thus supporting the RC assumption. However, this approach hinges on a participants' ability to quantify the behaviour in question and this is not always possible in a health domain.

Another approach is set out by Bago d'Uva et al. (2011) which requires an objective measure or a set of proxy indicators sufficient to soak up any residual covariation between the covariates (e.g. gender, education) and the construct being assessed. The objective measure or set of indicators is used to determine the thresholds respondents use to report on their own situation. These thresholds are then compared to the thresholds used by respondents to rate the vignettes. If RC holds then the thresholds should show congruence. Using this test, the assumption of RC was rejected for a set of vignettes used in the English Longitudinal Study of Aging (ELSA). However, when Grol-Prokopczyk et al. (2015) applied the same test they found only minor violations of RC for specific countries, while most countries showed good congruence, indicating the presence of RC.

In a more recent example, Dasgupta (2018) performed a test of similar test of RC. SRH was regressed onto a set of covariates after controlling for a set of proxy indicators. As the effects of "true" health are controlled for, any remaining systematic variation in SRH, as revealed by the covariate coefficients, can be attributed to systematic bias in reporting behaviour. The pattern of reporting heterogeneity for SRH revealed in the covariates is then compared to the reporting heterogeneity revealed in the same covariates for the AV responses. RC is indicated by the same signs in the regression coefficients for covariates for SRH and AV responses. Using this method, Dasgupta (2018) reports that RC was present for AV responses across six Indian states.

4.6.5.2 Vignette equivalence.

Vignette equivalence is generally considered to be the most demanding AV measurement assumption but He et al. (2017) contend that it is often tacitly assumed. Many of the studies which have tested and report meeting this assumption, including the WHO vignettes used in numerous studies, relied on a test that involves checking that most participants correctly rank order vignettes in a series (Kang & Grol-Prokopczyk, 2020; King et al., 2004; Mojtabai, 2016; Murray, et al., 2003; Weiss & Roberts, 2018), or a slightly more demanding variant involving comparing patterns among non-normative rankings or differences in ranking consistencies across national other groups (Kristensen & Johansson, 2008; Rice, Robone, & Smith, 2011).

However, Grol-Prokopczyk et al. (2015) suggest that while rank ordering is necessary for VE, it may not be sufficient. Instead, they support the use of a more novel and stringent approach to VE testing first proposed by Bago d'Uva et al. (2011). This test is based on the observation that while there may be differences between participants in how they rate a

single vignette, the perceived distance (along the latent spectrum) between any two vignettes in a series should be constant across groups. If the difference in severity rating between vignette 1 and vignette 2 is different for groups A and B, then this is indicative that there are fundamental differences in how the two groups perceive the severity of the vignette. However, if the difference score is the same then groups A and B may have placed the vignettes at different points on the severity scale due to differences in how they apply response thresholds but it can assumed that they understand fundamental severity of the vignette at the same level.

Grol-Prokopczk et al. (2015) report that using this test, existing WHO vignettes used in many of the articles already discussed routinely fail to meet the assumption of VE. Similar findings were reported by Bago d'Uva et al. (2011) for vignettes used by ELSA. However, while these results do undermine the AV method, they are not enough to dismiss it. The authors concede that this test and the RC test described above are very demanding. Additionally, the RC test requires some form of objective indicator of construct in question, and this is often not available due to resource constraints or the abstract nature of the construct under investigation. Grol-Prokopczk and colleagues also point out that the vignettes they tested were from a sample of 10 countries selected to be as diverse as possible. Some subsets showed only minimal violations of VE, and some, such as specific 2-country pairings, do not violate it at all. This is consistent with Corrado and Weeks (2010) who found violations of VE across 11 countries as a whole but not in certain small subsets of countries.

Grol-Prokopczyk et al. (2015) suggest that one of the main reasons for the violation of the VE assumption is that researchers have paid too much attention to linguistic and grammatical rules when translating the vignettes, and not enough to cultural norms and social practices. While the WHO vignettes are grammatically correct, they *"seem in numerous ways to invite different interpretations across national, religious, and/or socioeconomic groups"* (p.1724). The authors recommend that future researchers should aim for ***functional equivalence*** rather than just linguistic equivalence, and that this may require deviating substantially from the source text. They suggest that researchers need to collaborate with experts in translation and local culture to generate vignettes that achieve semantic, conceptual and technical equivalence across groups.

4.7 Summary of methodological issues in cross-cultural research

Cross-cultural research requires a considered approach that is sensitive to the challenges of evaluating invariance and minimising bias in measurement comparisons. There are numerous threats to the invariance of a scale, including construct, method, and item bias. These are commonly addressed in the literature and can, on the whole, be effectively combatted through careful translation and planning. Another threat to invariance however, which can be far more subtle and requires a more tailored and labour-intensive approach, is response category bias. This bias occurs where group differences in internal and external reference points cause systematic differences in how individuals interpret and utilise response thresholds.

4.8 Combined approach of current study.

The current project aims combine two approaches to enhance the cross-cultural validity of direct comparisons between India and the UK using the EPDS. Firstly, measurement invariance testing will be used to successively test each level of invariance required for valid cross-cultural comparisons. This will provide good general insight as to whether the measure is functioning equivalently between the two cultures, at a construct, metric and scalar level. This information will be used to complement the second aim, which is to use the anchoring vignette methodology to detect and adjust for DIF in how participants from each country are using the EPDS response scales. As AVs have not previously been used with the EPDS, chapter 5 will outline the development of the vignettes themselves and the method used to administer them in each setting. Following this, results from the measurement invariance testing and AV analysis will be presented in Chapter 6.

Chapter 5: Development of a set of Anchoring Vignettes for the Edinburgh Postnatal Depression Scale.

5.1 Introduction

The key aim in developing anchoring vignettes is to achieve response consistency and vignette equivalence (King, Murray, Salomon & Tandon, 2004). This can be achieved through a methodical but flexible approach to development and administration. As such, the following section will draw together and provide a synthesis of the guidance that has emerged organically regarding the development, translation, and administration of anchoring vignettes as the literature has developed. Following on from this, a detailed account of the development, translation and piloting of the anchoring vignettes in the current study is provided.

5.2 Anchoring vignette development in the literature

5.2.1 Vignette equivalence

There are three key recommendations to improving vignette equivalence (VE) when using AVs in health research. Firstly, in order to leave nothing to the imagination or subjective interpretation of the participant, vignettes should be written as objectively, clearly and concretely as possible, with reference to specific behaviours that can or cannot be done, and their frequency of occurrence (King et al., 2004). Secondly, vignette authors should be sensitive to the cultural norms and practices that characterise the target populations, ideally purging the vignettes of any references that might carry a different weight or meaning to a specific group. Thirdly, during the translation process, researchers should strive to produce content that is functionally and conceptually equivalent between groups (Bago d'Uva, Lindeboom, O'Donnell, & van Doorslaer, 2011; Grol-Prokopczyk, Verdes-Tennant, McEniry, & Ispany,2015).

However, although there is a call for specificity, caution is required in the level of specificity used. More recent research has suggested that being *overly* specific in vignette descriptions can actually lead to violations of vignette equivalence. Su, Willis, and Salomon (2017) found that using concrete, objective distances in vignettes describing visual acuity led to differing interpretations of the underlying construct as many participants did not have the ability to accurately estimate distance. Grol-Prokopczyk (2018) also found that vignette series which mentioned specific diseases that carry distinct cultural weights or relevance (e.g., heart disease, diabetes) can provoke more substantial violations of VE. Vignettes that describe health in general terms that are universally experienced, such as fatigue and pain appear to

invite fewer violations of VE than references to specific diseases. The authors conclude that *"universality should trump specificity and concreteness as a priority in vignette design"* (p.9).

Concerns regarding references to specific diseases were also raised by Vonkova, Bendle and Papajoanu (2017) who noted that the VE assumption is particularly challenged where the construct being assessed is multidimensional in nature. For example, depression is a multi-faceted disorder but is often assessed in large scale studies by the question "How much of a problem do you have with feeling sad or depressed?" (Mojtabai, 2016; Molina, 2017). The same question applied to a vignette could conjure up several dimensions or symptoms of depression against which to rate the character. Violations could then occur if each vignette does not adequately describe every dimension of depression and participants are left to interpret the relative weight and importance of unspoken symptoms. This may have a larger effect where one group has a more detailed understanding of a construct, as this knowledge is likely to create more variation in their interpretation. One solution to this dilemma is to break the overall construct down into individual and unambiguous symptoms for both the self-assessment and vignette components. This allows for comprehensive coverage of a multidimensional construct without the need for long vignettes. It also addresses the earlier encouragement to avoid references to specific diseases that may carry different cultural weights and instead focus on universally experienced symptoms. In this sense, the EPDS may be particularly suited to this methodology, as each of its 10 items focuses on a specific symptom of postnatal depression.

Thus, vignette wording needs to find a balance between generality and specificity. If it is too vague or if pertinent elements of the construct have been omitted, then respondents are likely to interpret it differently because there is too much ambiguity. However, if vignettes are too specific violations of VE can be incurred either because not all participants are able to make accurate mental representations for certain concepts, or because specific diseases, such as depression, may carry different weights across cultures. It is better to break down abstract constructs into concrete symptoms and to describe characters' health in general terms that can be explicitly defined, such as pain, laughter, and crying.

5.2.2 Response consistency.

Much of the guidance on developing vignettes which meet the response consistency (RC) assumptions comes from a key study by Au and Lorgelly (2014) who carried out an in-depth quantitative and qualitative investigation of different approaches to administration.

Qualitative interviews showed that participants were more likely to rate themselves and the vignette characters on the same scale if they imagined themselves in the same situation. This highlights two important recommendations. The first is to write the vignettes in such a way as it will be easy for participants to put themselves in the shoes of the character. This can be achieved by ensuring that the language and contextual cues used are relevant and natural to each target population. Content that is more relevant to one group should be avoided. Secondly, participants should be given explicit instructions to consider the vignette character to be of the same age and background as themselves, and to imagine themselves in that person's situation. An additional finding that participants tended to forget to do this indicates that this instruction should be repeated regularly throughout the administration. This approach was supported by Knott, Lorgelly, Black, and Hollingsworth (2017a) who found that RC was achieved after implementing these recommendations even when using formal, rigorous tests.

Au and Lorgelly (2014) also found that using vignettes that presented an overall health state, including 5 different dimensions of health, led to an improvement in RC compared to when they used vignettes that described a single dimension of health. It may be that a description of health that contains multiple symptoms appears to be more realistic, as poor health is very rarely confined to one specific symptom, and so participants find it easier to rate the character in the same way as they rate themselves. However, a downside of this approach is that vignettes are naturally longer, meaning that retaining all the information is more difficult and that different participants may pay more attention to certain parts of the vignette. This style of vignette may also introduce ambiguity in that some participants may consider each symptom independently while others may perceive individual symptoms in the context of the other symptoms described, and therefore rate each symptom as more severe.

Another suggestion is to switch the question order so that the AVs are presented before the self-rating questions, thus priming respondents to define the response scale in a common way (Hopkins & King, 2010). However, Knott, Black, Hollingsworth, and Lorgelly (2017b) argue against this approach as it may also contaminate answers to the original measure, making the results incompatible with results from the same measure used without being primed by vignettes. Other recommendations include that the placement of the self-assessment should be varied randomly, that the vignettes should be presented randomly and that the age and gender of the vignette characters should be matched to participants (Buckley, 2010, Knott et al., 2017a).

5.2.3 A trade-off between assumptions?

Knott et al. (2017a) speculated that the measures to improve RC may have had an adverse effect on VE. They went as far as to say that the jury was still out on whether it was possible to satisfy both RC and VE at once as they appeared to trade off against each other to some degree. They hypothesised that by inviting the respondent to interpret the character to be more like themselves, the perceived level of underlying health in the vignettes may become increasingly heterogenous. Likewise, while writing comprehensive, concrete, objective vignettes may minimise violations of VE by reducing potential ambiguity in vignette interpretations, it may also make it more difficult for respondents to use consistent thresholds across self- and vignette-ratings. Knott et al. (2017b) warned that researchers should be mindful that, *"recommendations put forward with regards to either one of the assumptions, may come at the expense of the other"* (p.399).

Thus, the ideal vignette should be written in such a way that it is specific enough to avoid ambiguity in interpretation but general enough that the respondent is able to imagine themselves in the same situation and apply consistent standards across their ratings of self and vignette characters.

5.3 Anchoring vignette development in the current study.

Methodological considerations regarding anchoring vignette development are split into three key areas: Development in English, Translation into Kannada, and Piloting. In turn, the approach taken in each of these three areas was informed by a number of practical considerations, including the number of vignettes required for the overall scale and each item and how the final task was to be administered. Thus, the broad rational guiding the development of the vignettes is provided first, followed by a detailed description of the three key areas.

5.3.1 Rationale: practical considerations regarding vignette development

The current approach was guided by previous findings in the literature and by practical considerations for the administering the vignettes in a low-literacy urban sample from Bangalore, India. Further, as this method had not been used with this measure or in this context before, a degree of flexibility was required and utilised in the development and administration of the vignettes.

As already explained, the anchoring vignette method deals with systematic bias in how individuals from different groups or cultures use response scales in self-reports. Thus, as the method is assessing DIF in the response scale and not the items themselves, where a

questionnaire uses a uniform response scale, there is no need to provide vignettes for every item. However, as the Edinburgh Postnatal Depression Scale (EPDS) has a different response scale for each of its ten items, a distinct set of vignettes was required for each item (see Table 5.3.1 for details of response scales). The initial plan was to administer the vignettes using an adaptive technique, where the two vignettes that represent the severity either side of the self-rating are selected and presented to participants (King, 2009). So, rather than each participant rating a range of vignettes representing each and every different level of symptom severity possible, each participant is only presented with two vignettes that theoretically reflect symptom severity either side of their own self-rating given on the EPDS when they completed it. The choice of vignette presentation is therefore adapted to reflect an individual's self-rating. In order to do this, six vignettes were required for each set of four response options (see figure 5.3.1). This approach was selected as it is presented as an effective way of saving time and participant burden while also increasing the quality and precision of the data obtained (King & Wand, 2006). Thus, a total of 60 vignettes were written, with each vignette describing a single depressive symptom in as concise a manner as possible.

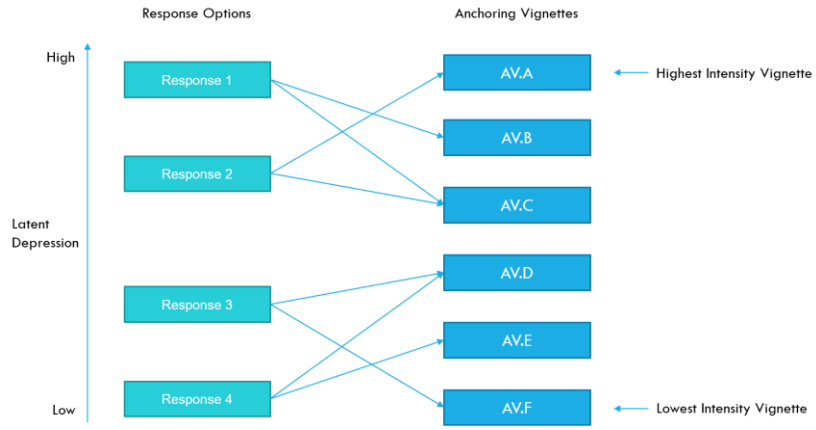
Table 5.3.1: Comparison of English and Kannada (back-translated) versions of the EPDS used in the current study.

	Item	Original English EPDS	Back-translation of Kannada EPDS
1	Stem	I have been able to laugh and see the funny side of things	No matter what I encounter with, I am able to find the funny aspect in that situation
	Response Set	As much as I always could	Just how much I could do before
		Not quite so much now	Not quite so much now
		Definitely not so much now	Definitely not as much as before
		Not at all	Not at all
2	Stem	I have looked forward with enjoyment to things.	I look forward to (been eager to) enjoying everyday activities of life
	Response Set	As much as I ever did	As much as I used to before
		Rather less than I used to	Rather less than I used to
		Definitely less than I used to	Definitely less than before
		Hardly at all	Hardly ever
3	Stem	I have blamed myself unnecessarily when things went wrong	I have blamed myself when the situation went wrong.
	Response Set	Yes, most of the time	Yes, most of the times
		Yes, some of the time	Yes, sometimes
		Not very often	No, once a while
		No, never	No, never
4	Stem	I have been anxious or worried for no good reason	I get anxious and worried for small matters
	Response Set	No, not at all	Not at all
		Hardly ever	Once a while
		Yes, sometimes	Yes, sometimes
		Yes, very often	Yes, most of the times

Table 5.3.1 continued: Comparison of English and Kannada (back-translated) versions of the EPDS used in the current study.

5	Stem	I have felt scared or panicky for no very good reason.	I get scared or panicky for small matters
	Response Set	Yes, quite a lot	Yes, most of the times
		Yes, sometimes	Yes, sometimes
		No, not much	Not much
		No, not at all	Not at all
6	Stem	Things have been getting on top of me	It is becoming impossible for me to cope with stress from various tasks
	Response Set	Yes, most of the time I haven't been able to cope at all	Yes, most of the times it gets difficult for me to handle
		Yes, sometimes I haven't been coping as well as usual	Yes, sometimes I have not been able to manage it as usual
		No, most of the time I have coped quite well	No, most of the times I have been able to manage
		No, I have been coping as well as ever	No, I am handling it as I have done always
7	Stem	I have been so unhappy that I have had difficulty sleeping.	I find it difficult to sleep, as I have feel unhappy
	Response Set	Yes, most of the time	Yes, most of the times
		Yes, sometimes	Yes, sometimes
		Not very often	No, once a while
		No, not at all	No, never
8	Stem	I have felt sad or miserable.	I have been sad or depressed
	Response Set	Yes, most of the time	Yes, most of the time
		Yes, quite often	Yes, often
		Not very often	No, once a while
		No, not at all	No, never
9	Stem	I have been so unhappy that I have been crying.	As I feel sad, I have been crying
	Response Set	Yes, most of the time	Yes, quite often
		Yes, quite often	Yes, sometimes
		Only occasionally	Once a while
		No, never	Not at all
10	Stem	The thought of harming myself has occurred to me.	I have had thoughts of harming or hurting myself
	Response Set	Yes, quite often	Yes, most of the times
		Sometimes	Sometimes
		Hardly ever	Rarely
		Never	Not at all

Figure 5.3.1: Showing adaptive selection approach to AV presentation



5.3.2 Task format and presentation.

While VE is largely affected by the content and translation of the vignettes, the likelihood of meeting the RC assumption rests more heavily on how the vignettes are presented and the clarity of the instructions (Au & Lorgelly, 2014). Hence, procedure was carefully considered during the development stage, with due consideration given to established methodology and the context in which the vignettes were administered. In the light of Au and Lorgelly's (2014) recommendation to repeat instructions throughout the task, but also in an effort to strike a balance between meeting this assumption and not overly frustrating participants, it was decided to repeat task instructions after every three vignettes.

We agreed with Knott et al. (2017a) that priming the participants with vignette completion prior to completing the self-assessment scale may contaminate answers to the latter and compromise future comparisons with other studies using the EPDS. Additionally, in the UK WCHADS cohort the EPDS had already been administered in the normal way without any AV priming, so a parallel procedure needed to be adopted in India to ensure responses to that scale were directly comparable and could just be adjusted for DIF following the results of AV analysis.

Finally, the order in which the vignettes were presented was considered. Buckley (2008) found that presenting the vignettes in descending order (of severity) attenuated vignette ratings and recommended that vignettes should be ordered randomly to mitigate context effects. This approach has been broadly accepted by researchers working in this field and general practice is randomise vignette presentation. Therefore, vignette presentation in

this study was randomised across items and severities, and placed after the self-assessment.

5.3.3 Development in English.

As the AV approach has not been used with the EPDS previously, all vignettes were developed by the author. Prior to writing the vignettes, the author collated and reviewed all existing vignettes that were available. The length and content of these vignettes were used as the starting point for the writing process. Vignettes were written in accordance with recommendations given in the literature that they should describe the exact level of the depressive symptom for each corresponding EPDS item, with reference to specific behaviours and emotions, and the frequency with which they occurred. Any reference to the overall concept of depression was avoided. Any language used in the response items (e.g. 'most of the time', 'quite often', 'definitely less' etc.) for the corresponding item was also omitted from the vignette so as not to bias a participant toward a particular response. Vignettes were written concisely using simple language and short sentences where possible. This was to ensure that all participants would be able to clearly understand and process the content. Vignettes in this study were presented orally and so it was important to keep memory load to a minimum. Therefore, each vignette description focused on a single depressive symptom corresponding to a single EPDS item.

Further to the description of the depressive symptom, each vignette contained a small amount of contextual information. This information was provided to make the vignette more realistic and to make the vignette character appear similar to a typical member of each target population. Additionally, certain items in the EPDS (1, 2 & 6) have response sets in which the participant must consider their current state (in the past 7 days) in relation to a previous state. This means that these vignettes were required to demonstrate both a previous and a current state of being.

Once written, the vignettes went through a stringent review and revision process. Feedback was initially provided by a group of experts from the UK and India on the cross-cultural appropriateness of the language used and appropriate revisions were made. Following this, to ensure the vignettes were appropriate for UK participants, additional feedback was provided by team members from the Wirral Child Health and Development Study regarding the reading level. Modifications included shortening sentences and replacing words that were deemed to be too complex. At this point the vignettes were also reviewed by expert clinicians and researchers in India and the UK to ensure they reflected the full range of

depressive symptoms assessed by the EPDS. Following this the vignettes were processed through a readability program to explore the reading age of the vignettes collectively and individually (https://www.webfx.com/tools/read-able/). As it was planned to administer the English vignettes to a sample broadly equivalent to the WCHADS sample in the UK, the mean reading age for the WCHADS sample (estimated to be 9 years) was used as a standard. Thus, the vignettes were modified to bring them as close to this reading age as possible. Due to the complex nature of some of the response sets, it was only possible to reduce the overall reading age of the vignettes to 13-14 years. The readability for vignette sets (one set corresponding to one EPDS item) ranged from 12-13 years to 14-15 years. Once the vignettes had been simplified to the highest possible degree, while still retaining their validity as descriptions of depressive symptoms, they were signed off as ready for translation by the author and the expert review panel. This process was not repeated for the Kannada vignettes, as it was planned to administer them orally, meaning reading age was not relevant.

5.3.4 Translation of vignettes into Kannada.

A major criticism of previous approaches is that studies have not taken adequate care in translating the vignettes so that they mean the same thing in different cultures. As stated earlier, the prevailing evidence and opinion from the literature strongly recommends a functional approach to translation. Thus, this project placed a heavy emphasis on attaining "functional equivalence" that prioritised equivalence of meaning over equivalence of wording.

To ensure a high level of functional equivalence and accuracy, the current approach to translation was based on World Health Organisation guidelines (Menon, Cherkil, Aswathy, Unnikrishnan, & Rajani, 2010). These guidelines set out the following steps: forward-translation, back-translation, and synthesis. A 'committee' approach was taken at each stage to further minimise bias, as recommended by Van der Vijver and Tanzer (2004). Further, although the English versions of the vignettes were developed first, a flexible approach was retained to allow for decentering and modification of both sets simultaneously (Erkut, 2010).

5.3.4.1 Forward translation.

Bilingual Panel. Vignettes were initially translated into Kannada by a team of 6 research assistants (RAs) from the Bangalore Child Health and Development Study. The RAs were divided into three pairs, and each pair translated 20 vignettes. All translators were native

Kannada speakers and fluent in English. Prior to translation, the team received training regarding the anchoring vignette method and instructions regarding translation. The RAs were instructed to focus on attaining functional equivalence in 3 areas: linguistic rules, cultural norms, and social practices (Pan & Fond, 2014). Linguistic rules refer to sentence structure, word order and grammar. Cultural norms refer to *how* people do or express things in the target culture. Social practices refer to *what* people do in a given culture. The team was given the freedom to deviate substantially from the original wording of the vignette if that would bring about a higher level of functional equivalence. Overall, it was emphasised that the translated vignettes should mean the same thing as the source vignettes and that all content should make sense to the target population.

Expert Panel Review. Following the initial forward-translation, the vignettes were reviewed by the author and the translation team in conjunction with the principal investigator of BCHADS, who is bilingual and a practicing psychiatrist. The first aim of the review was to ensure that the language used in the Kannada vignettes was appropriate for the target population. For instance, there were words or phrases which were technically correct but that were rarely used in conversation. In these cases, alternative wording was discussed by the team until a consensus was reached. The second objective was to ensure that the translations were conceptually equivalent to the source vignettes. The author was able to provide the context and meaning of phrases that the translation team struggled to understand. The author was also able to point out where there had been any clear misunderstandings of wording intent during the translation process.

5.3.4.2 Population review & back-translation.

Population Review. The revised vignettes were then taken out into the field for informal feedback from members of the target population. RAs were instructed to ask for feedback on understanding of individual words and phrases, as well the whole vignette. Vignettes were split into 6 sets of 10 with RAs instructed to ask for feedback from 2 participants for each set. Feedback from the population review was included in the final review and synthesis of feedback.

Back-Translation. Simultaneously, the Kannada vignettes were back-translated by a native speaking clinical psychologist who was also fluent in English and who had not taken part in the initial translation. This is an important part of the process as it allowed for the identification of discrepancies between the English and Kannada versions of the vignettes.

This individual was also asked to provide any feedback on conceptual equivalence and the appropriateness of the Kannada vocabulary used.

5.3.4.3 Final review and synthesis of feedback.

The final step of the translation process was to draw together and synthesise the back-translations and clinical psychologist feedback with the population feedback. The author met with the lead investigator of the BCHADS study, a bilingual psychiatrist and three members of the BCHADS research team who were all native Kannada speakers and fluent in English. The population review provided a better understanding of which words were not understood by the target population and so needed to be changed. Again, decisions were made by consensus within the panel. Final changes were then made to the Kannada vignettes based on any disparities in conceptual equivalence revealed by back-translations. Where it was not possible to achieve conceptual equivalence by changing the wording of the Kannada vignette (e.g., where there was no equivalent Kannada word to the English), or where it was easier to do so, changes were made to the source vignette. An example of the vignettes developed for EPDS item 1 is presented in Table 5.3.4.3, showing the English version that was initially translated into Kannada, the back-translation of the Kannada versions, and details of any changes made at this stage.

5.4 Pilot

5.4.1 Aim of the pilot study

Pilot testing forms a key part of developing new anchoring vignettes and there were several key areas that were explored during this part of the study. Of primary concern was whether the overall methodological approach was feasible in India and whether participants were able understand the vignettes. Understanding can be split into two areas. The first relates to semantic understanding. Did the participants understand the words and phrases used in the vignette, and does the overall vignette make sense to them? The second area relates to severity. Did participants understand the severity of the vignettes in the way that was expected (i.e., the *a priori* ranking)? Another key consideration was whether the vignette assumptions held for the vignettes. While it was beyond the scope of this study to perform stringent tests of assumptions, it was possible to examine various indicators of whether the assumptions were likely to be being met. A key paper that influenced the approach in this pilot study was Au and Lorgelly (2014), particularly with regards to the assessment of RC and participant experience of the task. As such, that study was used as a standard against which several of the indicators from this study were compared.

5.4.1.1 Research questions

1. Is the established anchoring vignette methodology a feasible and valid approach in India?
2. Do participants show semantic comprehension of the vignettes?
3. Are the vignette assumptions likely to be met in the current sample?

5.4.2 Method

5.4.2.1 Sample

The overall anchoring vignette assessment sat within the BCHADS cohort and was designed to be completed by a subset of participants. Of this subset, a convenience sample of Kannada-speaking mothers, who were also completing the wider battery of assessments, participated in the pilot study by completing the vignette assessment and providing additional feedback. A total of 32 mothers participated in the vignette assessment at this

Table 5.3.4.3: Example of anchoring vignettes for EPDS item 1, "I have been able to laugh and see the funny side of things", comparing the English version that was translated into Kannada and the subsequent back-translation, as well as any modifications made prior to piloting following final review.

	English Vignette (potential issues in red)	Back Translated Vignette	Kannada modifications made following review.
A.	X is usually very cheerful but recently she hasn't been feeling like herself. Situations that she would normally be able to make light of and brush off have been putting her in a bad mood. She used to laugh every day but she can't remember finding anything funny in the last week.	Lakshmi is usually very happy but recently she hasn't been feeling the same. Situations that wouldn't normally give her trouble have been now putting her in a bad mood. She used to be happy every day but she can't remember finding anything that makes her happy in the last week.	The 2 uses of "happy" were changed to words that matched English vignette.
B.	X is usually very happy but work has been getting her down recently and she has been much more serious than usual. In the last week she has probably only laughed once.	Sameena is usually very happy but recently due to work she has been much quieter than usual. In the last week she has probably laughed once.	Kannada vignette changed to reflect English text.
C.	X has always been a happy and **positive** person, but recently she has been feeling down about life. She would normally joke around and laugh with her friends every day, but she has only laughed a couple of times in the last week.	Meena has always been a happy and **positive** person, but recently she has been feeling low about life. She would normally joke and laugh with others every day, but in the last week she only laughed a couple of times	Hard to understand positive in Kannada (VS). Changed to "has a hopeful approach to life".
D.	X's mood has been very up and down this week. She has spent half the week feeling very low and barely able to force a smile. The rest of the time she has felt like her usual self and has been able to enjoy her time with friends and family.	Radha has been experiencing ups and downs in her mood this week. She was sad for almost half of the days of the week and found it hard to smile. At other times she felt like her usual self and has been able to enjoy her time with friends and family.	No changes made.
E.	X has been enjoying life just as she usually does for most of the week. There have been a couple of times where she has struggled to cope when things haven't gone to plan but otherwise she has been able to laugh off any difficulties.	Sheela enjoys her life through most of the week. When things don't go as expected she has struggled to cope on couple of occasions. Other than that even during difficult times she has been able to laugh.	No changes made.
F.	X is normally **very** happy in life and this week has been no different. While she has been busy with different jobs that needed doing, she has been able to stop and spend some time laughing and having fun with her children 3 or 4 times a day.	Shweta is normally happy in life and this week has also been the same. While she was doing all the activities that she was supposed to do, she has been able to also spend time with her children laughing and having fun/enjoyed doing so.	¾ times added to Kannada version.

point. Data from the pilot vignettes ratings were not included in any final anchoring vignette analyses.

5.4.2.2 Ethics

Ethical approval for the full BCHADS protocol, which included the use of the anchoring vignette methodology within the BCHADS sample, was given by NIMHANS and the University of Liverpool. Details of the dates of committee approvals for BCHADS are given in full in Chapter 3 (Section 3.2.1).

5.4.2.3 Measures

5.4.2.3.1 Maternal depression

Maternal depression was assessed in each cohort using the Edinburgh Postnatal Depression Scale (EPDS) (Cox, Holden, & Sagovsky, 1987), a 10-item Likert scale designed to detect depression in the postnatal period. For the purposes of the current study the EPDS was translated into the local language (Kannada) following World Health Organisation guidelines for measure translation. A comparison of the original English version of the EPDS and the back-translation of the Kannada version of the EPDS is presented in Table 5.3.1. Full details regarding the measure and translation process are provided in Chapter 3.

5.4.2.3.2 Anchoring vignettes

The anchoring vignettes were developed as part of this study for specific use with the EPDS. Each of the 10 EPDS items focuses on a specific symptom of postnatal depression and has its own unique response scale. A set of 6 vignettes was developed for each item. Each set of vignettes described characters that were exhibiting the symptom described by the corresponding EPDS item and that could be sensibly rated using the corresponding response scale. Within each set of vignettes, a range of symptom severities were portrayed, ranging from mild to severe. Therefore, a total of 60 anchoring vignettes were developed for use in the pilot. Full details of the development process are given in section 5.3.

5.4.2.3.3 Pilot feedback questionnaire

A questionnaire was developed to record feedback and provide additional data for evaluating the vignettes. This questionnaire was administered alongside the anchoring vignettes with provision being made for feedback to be given during the task and following the completion of the task, as described below.

Feasibility: Participants were asked to provide feedback on how they found the task overall by rating the vignettes on ease of understanding and the amount of concentration required, both on a scale of 1-10. The RAs administering the pilot were also instructed to observe and record how the mothers responded to the instructions.

Semantic Comprehension: Participants were instructed to highlight any words or phrases that they found difficult to understand at the outset of the questionnaire and were reminded and asked for feedback after each vignette. Any comprehension issues, including the number of times each vignette was listened to, were recorded and collated so that recurrent difficulties could be identified.

Response Consistency and Vignette Rating Behaviour: If instructions were not understood or effective, it would raise questions as to the feasibility of the AV method in this context (Au & Lorgelly, 2014; Knott et al., 2017a). Where findings suggested that this was the case, consideration was given to whether it was possible to make modifications to the instructions to enhance their effectiveness, without deviating substantially from what has been established in the literature. While Au and Lorgelly (2014) conducted semi-structured interviews with participants, that was beyond the scope of this pilot study. Instead, a short set of questions were designed to explore similar domains and completed by participants immediately following completion of the vignette task. This provided valuable insight into the thought processes involved in completing the vignette task and whether the likelihood that the response consistency assumption held. The response consistency assumption was explored by asking participants how much they agreed with the statement, "I rated my health on the same health scale (or in the same way) as I rated the health of the story characters". Responses were recorded on a 5-point scale (strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, and strongly disagree). Participants were also asked to describe their thoughts when rating the health of the vignette characters. Finally, participants were asked whether they assumed the characters were the same age and background as themselves, and whether they imagined themselves in their position. If the answer was negative, they were prompted to explain why.

5.4.2.4 Procedure

Prior to starting the vignette assessment, researchers introduced and explained the task to the participants using a pre-written script. Specific instructions to enhance response consistency (described in Section 5.2.2) were read out at the start of the task and then

repeated every three vignettes to ensure that they were not forgotten. Following the task introduction, participants completed a practice vignette.

The EPDS and AV assessment were presented on a tablet using the Qualtrics online survey platform. Due to low literacy levels the EPDS was researcher administrated. Each EPDS item was read out and participants were invited to select their response from the four response items. The AVs were pre-recorded and presented orally in a random order following the completion of the EPDS. Playback of the vignettes was carried out using headphones to ensure that mothers could hear them clearly.

A split-half approach to vignette selection was used, representing a part standard presentation approach and part adaptive approach, whereby Qualtrics was programmed to randomly present two sets of vignettes. Set A contained all the vignettes (6 per item) for EPDS items 1, 3, 5, 7, and 9 (n=30) and two adaptively selected vignettes for EPDS items 2, 4, 6, 8, and 10 (n=10). Set B contained the reverse. This ensured that (i) each vignette was presented at least 15 times while not extending the pilot such that it would place excessive demands on participants and (ii) each EPDS item received corresponding vignette ratings which theoretically represent all points in the response set.

Each participant was therefore presented with 40 vignettes in total. Each vignette was played once through the headphones. Following this, the RA would read out the response options and the vignette was played again so the participant could listen to it with the response options in mind. The RA would then re-read the options and the participant would select the response. All participants were encouraged to listen to each vignette at least twice and informed that they could listen to the vignette as many times beyond that as they wanted. Following each vignette, participants were asked if there was anything they struggled to understand or did not make sense. Any feedback was recorded by the RA. The number of times each vignette was played was also recorded.

Following the completion of the AV task, participants were asked a series of questions relating to the feasibility of the task and their thought-processes and approach to the task.

5.4.3 Analysis plan

Analysis used a combination of quantitative and qualitative approaches. Any quantitative analysis was carried out in SPSS 24. Analysis was performed in relation to each key area of interest as follows:

5.4.3.1 Task feasibility

Mean participant ratings for ease of understanding and the amount of concentration required were calculated and compared with Au and Lorgelly (2014). RA feedback was also considered.

5.4.3.2 Vignette comprehension (semantic)

Participant feedback was collated along with the number of times the vignettes were played. As most vignettes were played at least twice as part of the standard procedure, negative feedback was assumed where vignettes have been played 3 or more times. To determine the most troublesome vignettes, the two sources of data were combined and any vignette with 3 or more items of negative feedback was examined more closely.

5.4.3.3 Response consistency and vignette rating behaviour

Responses to the question, "I rated my health on the same health scale (or in the same way) as I rated the health of the story characters" were split into 3 categories regarding the likelihood of RC: Likely (strongly agree), Possible (somewhat agree, neither agree nor disagree), and Unlikely (somewhat disagree, strongly disagree). The proportion of responses for each category was calculated and compared against the responses reported by Au and Lorgelly (2014).

The proportion of respondents giving positive and negative responses to the items, "did you assume that they were of the same age and background as yourself?" and "did you imagine yourself in their position?" was calculated and compared against the findings from Au and Lorgelly (2014). The reasons given for negative responses were also examined.

Qualitative participant feedback regarding participant thought processes when rating the vignettes was examined in relation to their response to the items described above.

5.4.3.4 Vignette equivalence & comprehension (severity).

VE and severity comprehension were explored using a variation of the standard rank-order analysis used by King et al. (2004). The present approach was based on the King et al. (2004) approach by examining rank-order of the vignettes. If rank ordering was consistent between groups (e.g., countries) this was taken as evidence that the vignettes were evoking similar levels of the relevant construct in each population and the VE assumption was considered to be fulfilled.

Due to time constraints imposed by the timeline of the larger BCHADS study it was not feasible to pilot in both the UK and India simultaneously, meaning a direct comparison of

rank-ordering was not possible. Instead, the mean rank-ordering of vignettes in the Kannada sample was compared against the rank-ordering of a UK-based clinical psychologist. This individual was an expert in the area of maternal mental health and had worked extensively with the target population. It was therefore considered that she could offer a perspective on the rank ordering that was both clinically relevant and representative of the target population.

First, the mean response for each vignette across participants was calculated. There were four response items to select from for each vignette. These were scored from 0 to 3, with 0 representing the least severity of depressive symptoms and 3 representing the highest. These mean scores were then converted to ranks for each set of 6 vignettes. The rank order in the current sample for each set was then compared against the *a priori* ranking order and the ranking order of a UK clinical psychologist.

Some inconsistency in rank order was expected so it was important to distinguish between levels of inconsistencies (Murray et al., 2003). Where a vignette was two or more ranks out of expected position it was deemed to be a major violation. Where a vignette was one rank out of expected position it was deemed to be a minor violation. Where ranking inconsistencies were found the vignettes were reviewed to determine why this was the case and appropriate modifications were made.

5.4.4 Results.

5.4.4.1 Feasibility.

The mean rating for ease of understanding was 7.64 (SD = 2.43) and the mean rating for concentration required was 6.13 (SD = 3.20). Comparatively, the participants from Au and Lorgelly (2014) gave a mean rating of 8.3 out of 10 for ease and 7.4 out of 10 for concentration required.

5.4.4.2 Response consistency and vignette rating behaviour.

When asked how much they agreed with the statement, *"I rated my health on the same health scale (or in the same way) as I rated the health of the story characters",* 37% said they "Strongly Agreed", 56.3% said they "Somewhat" agreed, 3.1% said they "Neither Agreed Nor Disagreed" and 3.1% said they "Strongly Disagreed". Comparatively, Au and Lorgelly (2014) found that 38% of participants strongly agreed with the same statement.

100% of participants stated that they assumed the vignette characters were of the same age and background as themselves when rating the vignettes. 84.4% of participants stated they imagined themselves in the vignette scenario.

When participants were asked to describe their thought processes when rating the vignettes, most responded that they felt there was lots of similarity between their own lives and the situations and emotions described in the vignettes.

"What happens in my house every day, also happens in their lives"

“My family stories are also very similar. I could imagine myself in their place.”

Participants also reported imagining themselves in the place of the vignette character, trying to understand how difficult it was for characters, and thinking about what decisions they would make if they were in their place.

“While answering I imagine myself as a character from some stories.”

"I was trying to think what decisions I would take if I was in their place."

“I was trying to understand how difficult it is for them.”

Only 5 out of 32 participants indicated that they did not relate to the vignette characters or content.

5.4.4.3 Research assistant feedback.

The research assistants observed that a lot of the participants were not paying attention to the duration or frequency of the different emotions described in the vignettes, but rather were just looking for whether a feeling was there or not. In other words, if an emotion was mentioned, participants appeared to rate it without considering its longevity or intensity. This was first noted from observations that a number of participants were consistently rating vignettes as very severe when symptoms were mentioned but described as very mild or non-existent. Further inspection of the preliminary pilot data confirmed this pattern.

Research staff reported that it was difficult to ensure that participants were paying attention and listening properly to the vignettes due to the format in which they were presented. Participants would occasionally appear to disengage while listening to the vignettes.

Researchers reported that many of the participants did not want to listen to the vignettes twice when asked to and insisted that they had understood them the first time. This was supported by the records of how many times each vignette was played, with a substantial proportion of mothers only listening to each vignette once during the pilot.

5.4.4.4 Vignette comprehension.

6 of the 60 pilot vignettes received ≥ 3 negative feedback indicators (participant feedback and vignette play count) and were therefore implicated as troublesome vignettes (Table 5.4.4.4). A representation of the qualitative feedback is provided in Appendix 6.

Table 5.4.4.4: Vignettes with ≥ 3 items of indicators of poor comprehension

	Frequency of Indicators		
Vignette	**Participant Feedback**	**Play Count ≥3**	**Total Indicators**
10D	1	4	5
1E	2	2	4
1C	1	2	3
1D	3	0	3
2D	1	2	3
6F	2	1	3

5.4.4.5 Vignette equivalence and ranking comparison

Comparisons between the *a priori* rank order, UK clinician rank order and pilot rank order are presented in Tables 5.4.4.5a…j). Means, standard deviations and coefficient of variations are presented for each vignette. Ranking inconsistencies were distinguished as either major (2 or more ranks out of place) or minor (1 rank out of place). Violations are discussed and addressed in section 5.5.2.

Table 5.4.4.5a:
Ranking comparison of vignette ratings for EPDS Item 1: *"I have been able to laugh and see the funny side of things".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
1A	6	6	5*	1.5 (0.96, 0.64)
1B	5	5	4*	1.4 (0.9, 0.65)
1C	4	4	6**	1.56 (0.83, 0.57)
1D	3	3	2*	0.74 (0.79, 1.10)
1E	2	2	3*	0.83 (0.92, 1.10)
1F	1	1	1	0.25 (0.48, 1.79)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

Table 5.4.4.5b:
Ranking comparison of vignette ratings for EPDS Item 2: *"I have looked forward with enjoyment to things"*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
2A	6	6	5*	1.94 (0.97, 0.5)
2B	5	5	3**	1.78 (1.15, 0.63)
2C	4	4	2**	1.67 (1.15, 0.71)
2D	3	3	6**	1.97 (1.04, 0.53)
2E	2	2	4**	1.89 (1.21, 0.65)
2F	1	1	1	1.25 (0.92, 0.77)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

Table 5.4.4.5c:
Ranking comparison of vignette ratings for EPDS Item 3: *"I have blamed myself unnecessarily when things went wrong".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
3A	6	6	5*	1.69 (0.96, 0.56)
3B	5	5	3**	1.33 (0.9, 0.63)
3C	4	4	6**	1.84 (1.05, 0.55)
3D	3	3	4*	1.36 (0.65, 0.46)
3E	2	2	2	1.22 (0.65, 0.57)
3F	1	1	1	1.06 (1.00, 1.00)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

Table 5.4.4.5d:
Ranking comparison of vignette ratings for EPDS Item 4: *"I have been anxious or worried for no good reason".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
4A	6	6	5*	1.95 (0.88, 0.45)
4B	5	5	4*	1.65 (1.15, 0.71)
4C	4	4	6**	2.1 (0.85, 0.41)
4D	3	3	3	1.62 (0.79, 0.51)
4E	2	2	2	1.24 (1.14, 0.91)
4F	1	1	1	0.67 (0.99, 1.45)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

Table 5.4.4.5e:
Ranking comparison of vignette ratings for EPDS Item 5: *"I have felt scared or panicky for no very good reason".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
5A	6	6	5*	1.88 (0.8, 0.43)
5B	5	5	6*	2.06 (1.04, 0.5)
5C	4	4	3*	1.56 (0.74, 0.5)
5D	3	3	2*	1.31 (0.56, 0.41)
5E	2	2	4**	1.62 (0.91, 0.56)
5F	1	1	1	0.67 (0.77, 1.15)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

Table 5.4.4.5f:
Ranking comparison of vignette ratings for EPDS Item 6: *"Things have been getting on top of me".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
6A	6	6	6	1.94 (1.2, 0.62)
6B	5	5	4*	1.67 (1.28, 0.77)
6C	4	4	5*	1.83 (1.2, 0.65)
6D	3	3	2*	1.32 (1.14, 0.88)
6E	2	2	3*	1.48 (1.13, 0.76)
6F	1	1	1	0.58 (1.01, 1.76)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

Table 5.4.4.5g:
Ranking comparison of vignette ratings for EPDS Item 7: *"I have been so unhappy that I have had difficulty sleeping".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
7A	6	5***	5*	2.19 (0.95, 0.42)
7B	5	6***	6*	2.39 (0.88, 0.36)
7C	4	4	4	2.11 (1.02, 0.47)
7D	3	3	3	2 (1.03, 0.51)
7E	2	2	2	1.63 (0.99, 0.59)
7F	1	1	1	0.44 (0.88, 1.86)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.
***UK clinician rank inconsistent with a priori rank.

Table 5.4.4.5h:
Ranking comparison of vignette ratings for EPDS Item 8: *"I have felt sad or miserable".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
8A	6	6	6	1.94 (0.99, 0.51)
8B	5	5	2**	1.16 (0.97, 0.83)
8C	4	4	5*	1.6 (0.85, 0.59)
8D	3	3	4*	1.59 (0.81, 0.52)
8E	2	2	3*	1.24 (0.75, 0.58)
8F	1	1	1	0.38 (0.82, 2.11)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

Table 5.4.4.5i:
Ranking comparison of vignette ratings for EPDS Item 9: *"I have been so unhappy I have been crying".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
9A	6	6	6	2.07 (0.9, 0.43)
9B	5	5	4*	1.88 (0.86, 0.46)
9C	4	4	5*	2 (0.77, 0.4)
9D	3	3	3	1.73 (0.69, 0.43)
9E	2	2	2	1.48 (0.59, 0.43)
1F	1	1	1	0.25 (0.48, 1.79)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

Table 5.4.4.5j:
Ranking comparison of vignette ratings for EPDS Item 10: *"The thought of harming myself has occurred to me".*

Vignette	A Priori Rank	UK Clinician Rank	Pilot Rank	Mean Rating (SD, CoV)
10A	6	6	6	1.94 (0.96, 0.5)
10B	5	5	5	1.88 (0.98, 0.53)
10C	4	4	4	1.82 (0.86, 0.48)
10D	3	3	3	1.59 (0.9, 0.61)
10E	2	2	2	1.53 (0.79, 0.52)
10F	1	1	1	0.12 (0.34, 2.82)

Note: *Pilot rank inconsistent with a priori rank by 1 rank, ** Pilot rank inconsistent with a priori rank by ≥2 ranks.

5.5 Post-pilot review and modifications.

5.5.1 Review of vignettes using rank-order analysis.

Following the completion of the pilot data collection and analysis, an expert panel reviewed the data and discussed any changes that needed to be made to the vignettes. The panel consisted of the author, a clinical psychologist, a professor of biostatistics and psychometrics and a postdoctoral research assistant. All members of the panel have extensive experience in cross-cultural research. Following the agreed changes that were made by the panel, the modified vignettes were reviewed by the PI of BCHADS.

Modifications were made by examining the text in relation to rank order comparisons and vignette severity mean ratings. Several key issues were identified that could be driving rank order violations in the pilot data. The panel aimed to address these issues in the following ways. Firstly, to make any modifications that would enhance the conceptual equivalence of the vignettes. Secondly, to ensure that vignettes provide sufficiently distinct representations of symptom severity. Thirdly, to remove any possible primacy effects in the way vignettes are written. Data from the rank-order comparison was also cross-referenced with participant feedback.

A guiding principle of the review process was that changes were only made where absolutely necessary. Changes were only to be made where there was a clear rationale and where they were likely to result in significant contributions to meeting the aims described previously. Following modifications to the Kannada vignettes, the English vignettes were reviewed and compared to the Kannada vignette back-translations to determine if any changes could be made to further enhance vignette equivalence.

5.5.2 Modifications made to vignettes following rank-order review.

The following section provides details regarding the modifications made to the vignettes following examination of the rank-order data and a panel review of the vignettes to

determine any subsequent changes that needed to be made. Details are provided for each set of vignettes and an example of the changes are presented in Table 5.5.2.1. A complete record of the changes made to each vignette is provided in appendix 7.

5.5.2.1 EPDS item 1.

There was very little difference in the mean ratings of items 1A, 1B and 1C, suggesting that the vignettes did not provide the required specificity for distinguishing between response items. Examination of the text revealed several changes that could be made. Vignette 1C, rated as more severe than expected, described the character as a "positive person". This added an element of personality that vignettes 1A and 1B did not have. In turn, this may have increased the contrast between past and present states for vignette 1C, leading to higher severity ratings, and so the reference was removed. Participant feedback suggested that this vignette was difficult to understand, so the sentence order was changed after consultation with the BCHADS team. It was also decided that the reference to 'work' in 1B may have inferred that while things were difficult at work, things at home were okay, leading to lower ratings. The reference to work was therefore removed and replaced with a more general reference to "things".

The positions of vignettes 1D and 1E were switched compared to the expected rank order. Participant feedback suggested that both vignettes were difficult to understand, which may have led to the ranking violations. The vignettes were modified by removing some sections and switching some sentences around to make them easier to understand. It was also noted that the main cue in the EPDS item is to a positive behaviour, and that vignettes should follow this cue consistently by introducing the positive behaviour at the beginning of the vignette. Vignette 1D began with a negative behaviour and so was modified accordingly. See Table 5.5.2.1 for details of changes made.

5.5.2.2 EPDS item 2.

Mean ratings suggested that vignettes 2D and 2E were overrated in terms of severity so the panel focused on modifying these vignettes, rather than 2B and 2C. This EPDS item is also cued toward a positive emotion. Examination of the vignette text showed that vignettes 2D and 2E described the frequency of negative experiences, whereas the focus of the other vignettes in the set was on the frequency of positive experiences. This may have introduced primacy effects and provided a more negative picture for 2D and 2E, resulting in higher ratings. The focus and frequency of these vignettes was switched to follow the positive cue of the EPDS item. This also brought in a consistent format to the overall vignette set and

addressed feedback from participants that 2D was difficult to understand. Finally, the description of negative experience in vignette 2E was changed from “some time” to a “a short time”, as the former conveyed a higher level of intensity than was appropriate for this vignette.

5.5.2.3 EPDS item 3.

Mean ratings suggested that 3A was underrated and 3C was overrated. The wording of 3A was changed from “in a day” to “each day” to emphasise the frequency and severity of the symptom. The phrase “has a tendency to” was removed from vignette 3C as this may have inferred a trait-like or temperamental disposition toward self-blame, resulting in higher severity ratings for this vignette. Vignette 3B also appeared to be substantially underrated. This was addressed by moving the phrase “every day” to the beginning of the vignette to emphasise frequency. It was also suggested that the word “definitely” should be added to emphasise that she is blaming herself unnecessarily. However, after discussing this with the BCHADS team, it was discovered that there is no appropriate Kannada word for “definitely”. Colloquially, a word is dragged out and emphasised to communicate a sense of something “definitely” not being the mother’s fault. This colloquial emphasis was therefore added at the recording stage.

5.5.2.4 EPDS item 4.

Mean ratings suggested that the main issue in this set was that vignette 4C had been substantially overrated, and that if this was modified appropriately, the other ranking violations would correct themselves. It was decided to remove the word “crippling” from the description of anxiety in 4C.

5.5.2.5 EPDS item 5.

The main focus for this set was on reducing the intensity of vignette 5C as it showed the biggest violation. The word “very” was removed from the description of how fearful she was feeling. Furthermore, it was noted that this vignette did not reference a time period for how long it takes the character to recover from her fearful episode, unlike the other vignettes in this set. Thus, “for 20 minutes” was added to clarify the severity of this vignette relative to the others. Although, vignettes 5A and 5B were in the reverse position of their expected ranking, there was no clear rationale for how to change the vignettes. In this case, it was decided that no changes would be made as it would be creating extra work with no clear potential benefit.

Table 5.6.2.1: Showing post-pilot modifications to anchoring vignettes for EPDS item 1 ("I can laugh and see the funny side of things").

Back Translated Vignette (Changes following BT in red)	Pre-Pilot Changes	Post Pilot Modifications	Modification Rationale
Lakshmi is usually very happy but recently she hasn't been feeling the same. Situations that wouldn't normally give her trouble have been now putting her in a bad mood. She used to be happy every day but she can't remember finding anything that makes her happy in the last week.	The 2 uses of "happy" were changed to words that matched English vignette.	Lakshmi is usually very happy but recently she hasn't been feeling the same. Situations that wouldn't normally give her trouble have been now putting her in a bad mood. She used to be happy every day but she can't remember finding anything that makes her happy in the last week.	**NO CHANGES MADE FOLLOWING PILOTING**
Sameena is usually very happy but recently due to work she has been much quieter than usual. In the last week she has probably laughed once.	Kannada vignette changed to reflect English text.	Sameena is usually very happy but recently **things** have been getting her down and she has been much more serious than usual. In the last week she has probably laughed once.	Work changed to things as HS believed that this reference may infer that while things are difficult at work, things at home are ok, leading to lower rankings.
Meena has always been a happy and **positive** person, but recently she has been feeling low about life. She would normally joke and laugh with others every day, but in the last week she only laughed a couple of times	Hard to understand positive in Kannada (VS). Changed to "has a hopeful approach to life".	Meena has always been a **happy** person, but recently she has been feeling low about life. She would normally joke and laugh with others every day, but in the last week she only laughed a couple of times	Reference to "positive" or "hopeful" person removed as this introduces an added element of personality that A and B do not have. This may be increasing the contrast between past and present states and be leading to higher ratings.
Radha has been experiencing ups and downs in her mood this week. She was sad for almost half of the days of the week and found it hard to smile. At other times she felt like her usual self and has been able to enjoy her time with friends and family.		Radha felt like her usual self and has been able to enjoy time with her friends and family for 3 or 4 days this week. The other days she was sad and found it hard to smile.	**3 comments from participants saying it was too long/difficult to understand**. Removing first sentence should make it simpler to understand. Also change from almost half the days to 3 days. HS and AP also agreed that we should be consistent throughout the set in keeping the reference to positive behaviour as this is cued in the EPDS item.
Sheela enjoys her life through most of the week. When things don't go as expected she has struggled to cope on couple of occasions. Other than that even during difficult times she has been able to laugh.		Sheela enjoys her life through most of the week. **She has struggled to cope on a couple occasions when things didn't go as expected**. Other than that even during difficult times she has been able to laugh.	**Participant feedback suggests this is difficult to understand.** Switched sentence around to make it read more easily.
Shweta is normally happy in life and this week has also been the same. While she was doing all the activities that she was supposed to do, she has been able to also spend time with her children laughing and having fun/enjoyed doing so.	¾ times added to Kannada version.	Shweta is normally happy in life and this week has also been the same. While she was doing all the activities that she was supposed to do, she has been able to also spend time with her children laughing and having fun/enjoyed doing so.	**NO CHANGES MADE FOLLOWING PILOTING**

Note: Where changes were made to back-translated vignette prior to piloting they were made in the Kannada versions only and so will only be reflected in the final back-translations that were conducted following piloting and post-pilot modifications. Therefore, the Kannada version presented to pilot participants varied slightly from those presented above.

5.5.2.6 EPDS item 6.

On examining the text, it was decided to change the wording in 6B from “burdened” to “over-burdened” to increase the level of intensity relative to 6C. No changes were made to 6C. For 6D, which was underrated, the word “all” was added to emphasise a large amount of household responsibilities. The translation referred to looking after the family in a holistic way, rather than practical household responsibilities, so this was corrected. For vignette 6E, which was overrated, the wording was changed from “of late” to “this week” to minimise the timeframe of the difficulties. Although there was some feedback that 6F was too long and difficult to understand, the mean rating suggests that, generally, it was adequately understood and so no changes were made.

5.5.2.7 EPDS item 7.

The panel agreed with the UK clinician and pilot ranking of vignettes 7A and 7B. The rationale behind this thinking was that taking a long time to fall asleep (7B) was likely to be understood as more of a problem than waking up several times during the night (7A). Thus, no changes were made to the content of the vignettes, but the a priori ranking of 7A and 7B was switched.

5.5.2.8 EPDS item 8.

The panel focused on modifications to 8B as it was substantially underrated and looked to be driving the other minor ranking violations. The phrase “manages to get out of bed with difficulty” was identified as problematic. It was noted that due to adverse socio-economic conditions, many women in India may “manage with difficulty” a lot of the time, and therefore this would not be adjudged to be very severe. This was changed to “struggles to get herself out of bed every morning” in order to emphasise how difficult it is and that it is a daily occurrence. The first sentence, “feels very sad in the mornings” was also removed as it restricts feelings of sadness to the mornings. This change should open up the feelings of sadness to being pervasive throughout the day.

5.5.2.9 EPDS item 9.

The phrase “each night” was added to the description of sleep problems in 9B to emphasise that this is something that happens every day and increase its perceived severity relative to 9C.

5.5.2.10 EPDS item 10.

No changes were made to vignettes for EPDS item 10. Although there were 5 participant indicators that vignette 10D was troublesome, the comment from participants was that it

was difficult to answer. This suggests that the difficulty may lie in the response items themselves and not the vignette. It was decided that, because the ranking order was correct and there was no clear rationale for changes to be made, these vignettes would not be modified.

5.6 Discussion

5.6.1 Summary of findings

One of the main aims of the pilot study was to explore the feasibility of the AV method in India. Findings suggested that participants generally understood what was required of them and were able to complete the task. Some contextual limitations do exist, but in most cases minor procedural adjustments could be made to rectify these issues. Participants also generally demonstrated good semantic understanding of the vignette content. Thus, the AV approach appears to be valid and feasible in this context. Feedback regarding reporting behaviour indicates that the RC assumption is likely to have been met but rank-order comparisons revealed a number of inconsistencies between the expected and actual mean rank order. Further examination revealed that this was due to a number of reasons, including isolated cases of poor conceptual equivalence. These findings highlighted problem areas of the text which could then be modified appropriately.

5.6.2 Feasibility and administration.

Ratings regarding ease and concentration given in the current study were comparable to those found by Au and Lorgelly (2014). Although participants found the vignettes slightly less easy to understand, they also reported that slightly less concentration was required to evaluate the characters. Feasibility was further demonstrated by participant feedback that the situations described in the vignettes were relevant to their lives and that many of them had experienced similar emotions and scenarios. As the scenarios were devised by a UK resident and approved by various UK experts, it can be assumed that they are also relevant to the UK context, thereby indicating that it is possible to write vignettes that are relevant to both target populations.

Due to the low literacy rate of the BCHADS sample, vignettes in this study were pre-recorded and played to participants using headphones. This made it difficult for test administrators to know when a participant had disengaged. Another limitation is that it is difficult to retain the necessary information in this format, and a participant cannot glance back over the text for a reminder of the relevant details. To combat this, vignettes were written to be as short and concise as possible, thereby minimising cognitive load.

Additionally, the intended procedure was that participants would listen to each vignette at least twice. However, feedback suggested that many of the participants grew frustrated when being asked to listen to the vignettes twice and stated they did not need to. Due to the length and the details included in the vignettes, it was highly unlikely that participants were recollecting them accurately after listening to them once. Extra training was given to the test administrators prior to the full AV assessment so that they were able to provide better reasoning for why participants should listen to each vignette twice. However, where participants still did not want to listen twice, they were not made to do so. Therefore, this needs to be considered as a limitation of the AV method in this context.

An element that may have added to participant frustration or disengagement with the AV task was the length of the Kannada vignettes. All of the vignettes increased in length to some degree, some very significantly, when translated from English. Although every effort was made to make the vignettes as concise as possible, the nature of the EPDS and the requirements of the AV method restricted these efforts. In view of the difficulties caused by the length of the Kannada vignettes, it may have been wiser to sacrifice some of the contextual information. While this may have detracted from RC fulfilment, it may have increased participant engagement in the overall task and reduced violations of VE. This trade-off is something to consider for researchers using the AV method in this context in the future. Researchers should consider how vignettes might be elongated in the target language when developing the initial vignettes so that they can make a more informed decision regarding the optimal balance of context and vignette length.

5.6.3 Vignette assumptions.

Test administrators noted that participants did not seem to be paying attention to the frequency or duration of emotions and behaviours in the vignettes. Instead, participants were focusing on whether an emotion or behaviour was present or not. Without the frequency, there is a lot of room for individual interpretation, which may have led to violations of the VE assumption. As a result, it was decided to add a direct instruction to pay attention to the duration of the symptoms that characters experienced. This was inserted into the procedure part way through the pilot and researchers noted that participants were paying more attention to the frequency descriptors as a result.

The established procedures also have a major impact on RC. Both the content and the timing of the instructions given to participants have been carefully curated to increase the likelihood of meeting this assumption (Knott et al., 2017a). Thus, if results indicate that the

RC assumption is not being met, this could be indicative that the established procedure, developed in a western setting, does not function adequately in an Indian context. However, 37 % of participants stated that they "strongly" agreed with the statement that they rated the vignette character in the same way that they rated themselves, while 56.3% said they "somewhat" agreed. Au and Lorgelly (2014) gained similar results. They interpreted these responses as indicating that it was either "likely" or "possible" that the RC assumption was met, respectively, and reported that RC was likely to have been met for 38% of participants. Further to this in the current study, 100% of participants indicated that they followed instructions and considered the participant to be the same age and background as themselves, and 84.4% imagined themselves in the position of the character when rating the vignette.

These findings were supported by feedback from many participants that the vignettes contained situations and emotions that they had experienced in their lives (Appendix 6). With this being the case, participants will have been more easily able to empathise with the vignette characters and put themselves in their position. In turn, this increased the likelihood that participants rated the characters in the same way they rated themselves. While stringent tests of RC were beyond the scope of this study, Knott et al. (2017a) followed all the same recommendations and found that the RC assumption was met. Thus, it is reasonable to interpret the current findings as being indicative that the vignettes are likely to meet the RC assumption, and that the established procedure is effective in this regard.

The vignettes themselves were then considered and evaluated in detail. Using a rank-order comparison and participant feedback to draw attention to any possible issues, a panel of experts examined the text and made any necessary modifications. Rank order violations were found to arise for several reasons. Firstly, there were some inconsistencies between the focus of the vignettes and the cues given in the corresponding EPDS item. For example, item 2 reads *"I have looked forward with enjoyment to things".* While the majority of vignettes in this set followed the cue in the item and focused on the frequency and intensity of the positive emotion, there were two vignettes that focused on the frequency and intensity of an opposite negative emotion. It is possible that this resulted in a primacy effect that inflated perceptions of the severity of the depressive symptom. This issue was noted in several vignettes and corrected. Secondly, several of the vignettes contained unclear representations of symptom severity, such that it may have been difficult for participants to distinguish between vignettes when selecting response items. Some minor

modifications were therefore required to clarify the intensity of symptoms portrayed. Thirdly, several vignettes were not adequately conceptually equivalent. In most cases, this was due to translation issues that arose because a certain word or phrase had been translated incorrectly. For example, one vignette refers to "household responsibilities". This was meant to refer to practical jobs around the house, but when translated into Kannada, took on the meaning of looking after the family in a more holistic sense. In other cases, conceptual differences were driven by phrases that had the same meaning but may have carried different cultural weights. For example, a description of someone who "manages with difficulty" may appear less remarkable and severe to an Indian participant due to increased socioeconomic adversity. Finally, there was one instance where the UK clinician and pilot rank order both disagreed with the *a priori* rank order. In this case, the *a priori* rank order was altered.

This process of review and modification was a vital part of the wider vignette development process. Beyond enhancing conceptual equivalence, it was also important that the vignettes were distinct enough that they clearly represented discrete levels of the relevant depressive symptom. If vignettes were not sufficiently distinct they would not provide a clear or accurate indication of DIF. Following the changes that were made to the vignettes in the post-pilot phase, the expert panel was satisfied that this was the case.

5.6.4 Limitations.

Although every effort was made to enhance the validity of the vignettes, there were certain exogenous factors that could not be addressed by vignette design. Two of these issues arose from the EPDS itself. The phrasing of the items is complex and the BCHADS cohort has experienced difficulties in understanding and responding to them. Every effort was made in the initial translation of the EPDS to phrase the items in a way that was easily comprehensible to the study population but participants were still experiencing some difficulties. The second issue is that some of the EPDS response sets are more complex than others. For example, the first response option for item 7 is, *"Yes, sometimes",* whereas the first response option for item 6 is, *"Yes, sometimes I haven't been coping as well as usual"*. The option for item 6 clearly requires more concentration and places a heavier cognitive load on participants than the option for item 7. This increases the likelihood of variation in ratings for item 6 as participants face the dual challenge of interpreting the meaning of both the vignette and the response set. There were also some English words that did not have a Kannada equivalent.

5.7 Conclusion.

This chapter details how a set of anchoring vignettes were developed for use with the EPDS in India and the UK. The pilot provided an opportunity to further develop the vignettes by observing how participants understood and evaluated them. It also afforded a vital insight into how the AV method functioned in this context. Findings indicated that participants demonstrated an ability to understand and complete the vignette task. Results also showed that participants rated vignettes in the same way as themselves and that vignette content was relevant to participants' lives. There were some limitations introduced by the EPDS itself and by the use of the AV method in the contrasting UK and Indian contexts. Some of these limitations could be addressed through minor procedural modifications. Other limitations, such as the nature of the EPDS, were outside of the control of the research team but they did not critically undermine the approach. Overall, the AV method appears to function well in this context and, following appropriate adjustments, should be successful in meeting the aim of enhancing cross-cultural comparisons of postnatal depression using data drawn from the EPDS. The final set of anchoring vignettes were therefore deemed acceptable for use in this context and the full assessment was administered. An example of the final vignettes is provided in Table 5.7 and shows the final English and Kannada versions, as well as the back-translation of the final Kannada version (see Appendix 8 for complete set of final vignettes). The results of the empirical AV investigation in India and the UK are presented and discussed in the next chapter.

Table 5.7: Final English, Kannada and back-translated anchoring vignettes for EPDS item 1 (I have been able to laugh and see the funny side of things)

AV	English Final Version	Kannada Final Version	English Back-Translation
A	X is usually very cheerful but recently she has not been feeling like herself. Situations that normally wouldn't bother her have been putting her in a bad mood. She used to laugh every day, but she can't remember finding anything funny in the last week.	Lakshmi saadharanavaagi thumbaa khushiyaagi irutthaare. Aadhare munchinanthe anisuthilla. Maamuliyaagi avarige thondhare kodadha sandharbhagalu, eega avarannu ketta manasthithige tharutthidhe. Avaru dinaa nagunagutthaa irutthiddaru.aadhare, kaledha ondhu vaaradindha avarige yenu thamashe thandhu koduva haage jnyaapaka barutthilla.	Lakshmi will usually be very happy. But she's not feeling like before. The contexts/ situations which usually won't bother her are creating bad moods. She used be cheerful every day. But, From past one week, she can't remember things which give her joy (she can't remember any that will bring her fun).
B	X is usually very happy but work has been getting her down recently and she has been much more serious than usual. In the last week she has probably only laughed once.	Sameena saamaanyavaagi bahala santhoshadinda iruttare. Adare itthichege eladarinda kuggiddale haagu ghambeeravaagiddale. Kaleda ondu vaaradalli bahusha avaru onde baari nakkirabahudu.	Sameena will usually be very happy. But, she is in low spirits because of day to day challenges and is very serious now a day. Probably, she must have laughed only once this past week.
C	X has always been a happy and positive person, but recently she has been feeling down about life. She would normally joke around and laugh with her friends every day, but she has only laughed a couple of times in the last week.	Meena yaavagalu santhoshavaagi iruttaare. Aadhare itthichege avarige jeevanadalli kuggutthiruvanthe anubhava aagutthide. Avaru saamanyavaagi thamaashe maaduttha, yellarondhige prathidina nagunaguthaa iddharu, aadhare kaledha vaaradalli avaru kevala vondheradu baari nakkiddhaare.	Meena is happy all the time. But she's feeling low in life now a day. She usually used to make fun, used to be cheerful with everyone every day. But, she has laughed/ smiled one or two times in the last week.
D	X has been able to enjoy time with friends and family 3-4 days a week. The other half of the week she feels very low and barely able to force a smile.	Sitage thaanu munche iruvantheye anisuthade, haagu varadalli 3 athava 4 dinagalu, snehitharodane-kutumba davarodane anandisalu saadhyavaagutide. Ulida dinagalu avalu dukhadinda iddalu mattu nagalu saha kashta vagithu.	Sita feels that she has been the same as ever before, and 3 or 4 times in a week, she can enjoy the time with the friends and the family members. She was sad and had difficulty even to smile rest of the days.
E	X has been enjoying life just as she usually does for most of the week. There have been a couple of times where she has struggled to cope when things haven't gone to plan but otherwise she has been able to laugh off any difficulties.	Sheela Vaaradalli hechhu dinagalu santhoshavagi iddaare. Andhukondanthe kelasagalu nadeyadiddaaga adannu nibhayisikondu hogalu kelavomme kashta vaagutthittu. Adannu bittare kashtadallu avalu nagalu sadyavaayithu.	Sheela is happy most of the days in a week. At times had difficulty to manage when the work didn't happen in the expected way. Apart from that she could smile even in tough times.
F	X is normally very happy in life and this week has been no different. While she has been busy with different jobs that needed doing, she has been able to stop and spend some time laughing and having fun with her children 3 or 4 times a day.	Shwetha saamaanyavaagi santoshavaagi iruthare. Ee vaara kooda haage iddaru. Avaru maada bekaagiruva halavaaru kelasa karyagalunnu maadutta, kelavu samayavannu makkala jothege 3 rinda 4 baari moju mastiyalli kaleyuvaru.	Shwetha will usually be happy. She was feeling the same this week also. Along with daily chores which she is supposed to do, she spends time to have fun with children 3 or 4 times a day.

Chapter 6: Using anchoring vignettes to investigate and adjust for cultural differences in response style between India and the UK.

6.1 Introduction to aims of the Study

Building on the background literature summarised in Chapter 4 and the development of Anchoring Vignettes for the EPDS described in Chapter 5, this Chapter reports results from an empirical investigation of the cross-cultural functioning and validity of the Edinburgh Postnatal Depression Scale (EPDS) conducted using two urban community samples in India and the UK. There were two principal study aims which will be described in turn.

6.1.1 Aim 1 – Evaluation of measurement invariance in the EPDS

The first aim was to evaluate measurement invariance (i.e., configural, metric, and scalar) in direct comparisons between the two samples. Measurement invariance has been highlighted as a necessary but rarely studied condition for the valid comparison of EPDS scores across groups and time-points, with existing studies reporting inconsistent results (Cunningham, Brown, & Page, 2015; Di Florio et al., 2017). This study was designed to address this in two ways. First, longitudinal invariance was investigated within in each cultural setting by using longitudinal item factor analysis techniques (LIFA) to examine differences in reporting on the EPDS between the early and later postnatal periods. This is important because of the considerable biological, physiological and contextual changes that a mother experiences during the first postnatal year which may affect invariance (Martin & Redshaw, 2018). Second, cross-cultural invariance was tested using established multi-group item factor analysis (MIFA) techniques to investigate the influence of sociocultural factors unique to each setting on EPDS responses (Milfont & Fischer, 2010).

6.1.2 Aim 2 – Detection and minimisation of bias using the Anchoring Vignette methodology

The second aim, which is the main focus of this chapter, was to detect and minimise bias. To do this, a novel set of anchoring vignettes was used to detect and adjust for differential item functioning (DIF) in how participants from each group understand and apply response category thresholds (King, Murray, Salamon, & Tandon, 2004). This is the first study to use anchoring vignettes with the EPDS and the first to apply a parametric modelling approach to a multi-item scale in health-related research. Although, this method has previously been used to apply DIF adjustment to multi-item scales in personality research, those studies have typically used a non-parametric approach (He et al., 2017; Marksetiner, Kuger, & Kleime, 2019; Mottus et al., 2012). The main focus of the health literature, both parametric and non-parametric, so far has been the correction of single item measures (Kong & Grol-Prokopczyk, 2020; Knott et al., 2017; Mojtabai, 2016; Molina, 2017). As such,

a secondary aim of the study was to investigate the utility of different parametric approaches to anchoring vignettes in this context and with this scale. Detailed attention will be given to understanding how the methodology is working and implications for its future use in this context will be discussed.

Another novel aspect of this study is the application of a correction factor generated in one sample to the self-ratings generated in a different sample. Although King et al. (2004) do note that the parametric method can be used in this way, it rarely, if ever, is. Typically, the measurement scale and the anchoring vignettes are completed by the same participants. In this study, this was the case in India, but not in the UK. The UK perinatal cohort (WCHADS) was established several years earlier than the Indian cohort (BCHADS) and so the EPDS data at the relevant time-point had already been collected. Therefore, a new sample, termed the UK Anchoring Vignette sample (UKAV), was recruited which was matched to WCHADS UK participants on key eligibility criteria (maternal age, nationality, child age). The similarity between the samples will allow for the correction score to be generated in the UKAV sample and applied to the other (WCHADS). This is a novel approach to the AV method and the implications are that if a correction score can be generated in a representative sample, that correction score can then be applied to other cohorts with similar characteristics.

The current chapter begins with a summary of the study design and methods used, and a description of the two AV modelling approaches. Following this, results from the measurement invariance (MI) testing and the anchoring vignette (AV) analysis are presented and are discussed in terms of relevance to the existing literature and their implications for future cross-cultural research.

6.2 Methods

The current study reports data derived from two established cohorts (BCHADS and WCHADS) and a new sample recruited specifically for this project (UKAV). Each will be described in turn. Data collection points were selected to maximise the data available for comparison between BCHADS and WCHADS, but also to provide the most theoretical relevance, driven partially by the results of the EPDS validation study, in which distinct clinical thresholds were indicated at 8 weeks (≥3) and 6-24 months (≥10) in BCHADS (Section 3.2.2.1, Appendix 4).

6.2.1 Bangalore Child Health and Development Study (BCHADS) design and sample

BCHADS is a prospective epidemiological cohort study starting in pregnancy and designed to study early risk and protective factors for childhood mental health throughout pregnancy and infancy, in India. 909 expectant mothers living in urban slums in Bangalore were recruited from 3 community antenatal clinics when attending routine appointments. Following exclusions, 825 remained eligible

for postnatal follow-up. Detailed information regarding ethical approval, sampling strategy, attrition and assessment phases are provided in chapter 3 for the BCHADS sample.

EPDS data reported in this study was collected at 8-12 (T5) weeks and 12 months (T8) postpartum. 544 participants completed the EPDS at T5 and 549 participants completed the EPDS at T8.

As discussed previously, the anchoring vignettes were developed as part of the BCHADS project and inserted into the assessment protocol at the point when data collection for T8 was coming to an end and data collection for T9 was ongoing. Therefore, in order to maximise available data, the anchoring vignettes were administered at both T8 and T9 and combined into a single dataset. A total of 247 participants completed the AV task and were included in the AV subsample. Table 6.2.1 provides the key demographic details of the final sample of each phase. The samples are broadly similar but a lower proportion of participants in the AV subsample had completed up to secondary education.

Table 6.2.1: Maternal characteristics in the final BCHADS T5 and T8 samples and the AV subsample.

Characteristic		BCHADS T5	BCHADS T8	BCHADS AV Subsample
N		544	549	247
Mother Age	(M, SD)	23.16 (3.51)	23.24 (3.56)	23.37 (3.57)
Education	(up to secondary)	71.2%	70.1%	64.0%
Employed	Yes	14.7%	13.4%	15.0%
Partner Status	Married	100.0%	100.0%	100.0%

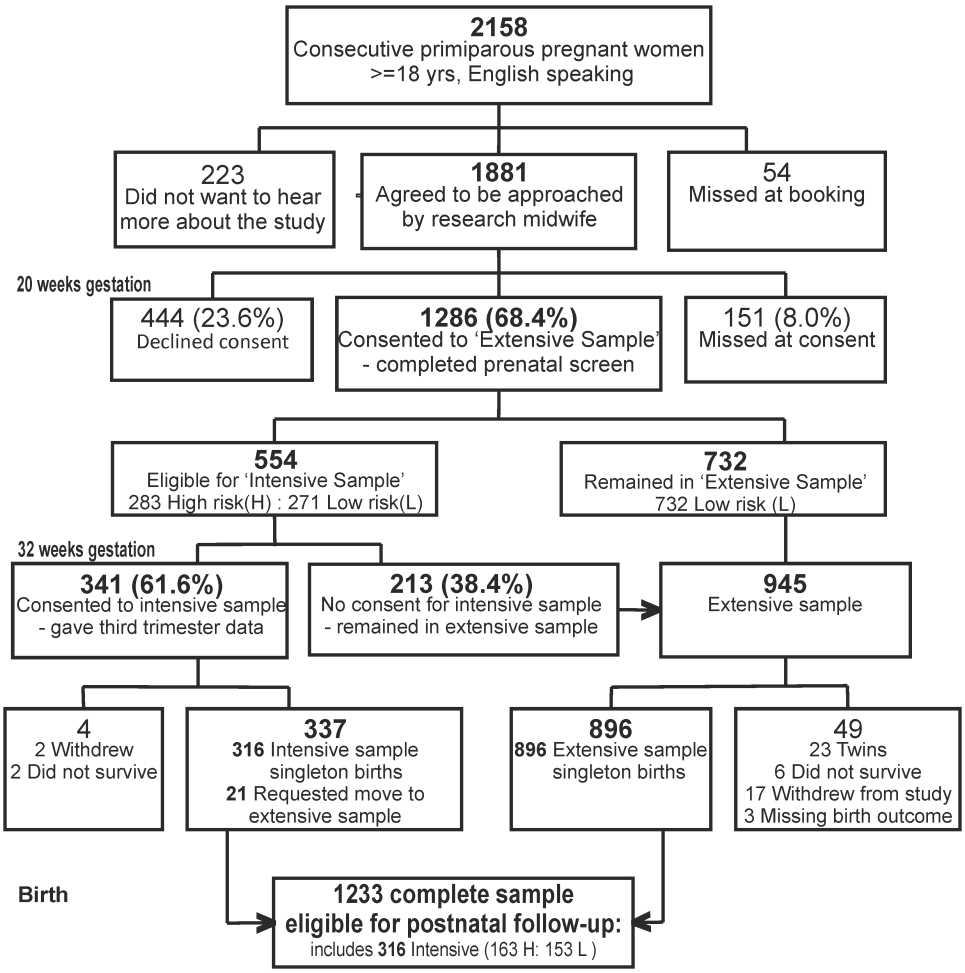
6.2.2 Wirral Child Health and Development Study (WCHADS) design and sample

The WCHADS is a prospective epidemiological cohort study starting in pregnancy designed to investigate the earliest origins of childhood conduct problems. The study used a two-stage stratified design in which a larger ('extensive') general population sample was used to provide a stratified random subsample ('intensive'), and both were then followed up in tandem over time. Only extensive sample data will be analysed in the current study.

Recruitment to WCHADS took place between 12th February 2007 and 29th October 2008. A consecutive sample of 2158 expectant mothers were approached at their 12-week antenatal appointment at an NHS hospital antenatal clinic and asked if they would like to hear more about the study at their 20-week scan. Women were invited to participate in the WCHADS if they were: (i)

primiparous, (ii) English speaking and (iii) 18 years of age or above at the time of recruitment. They were subsequently excluded if their baby was later found to have a gross congenital abnormality or did not survive. Multiple births were also excluded from further follow-up. No exclusions were made on the basis of premature birth (<37 weeks) or low birth weight (<2500g), or late registration for antenatal care, as these events have been associated with prenatal stress in previous research. Figure 6.2.2 shows the full recruitment process.

Figure 6.2.2: WCHADS Sampling Flow-Chart



6.2.2.1 Recruitment to extensive sample

After the study was introduced to eligible women by clinic midwives at 12-weeks, 1881 expressed interest in hearing more about the study. These women were approached by midwives at their 20-week scan and 1286 (68.4%) consented to take part in the study. Women seen at this point who did not consent (n= 444) were asked for basic demographic information.

6.2.2.2 Comparison of consenters and non-consenters to the extensive sample

Basic demographic data, comprising of age and postcode, were collected from all women who expressed interest in hearing more about the study when they were first approached by clinic midwives. Postcodes were used to determine socioeconomic status via the revised English Index of Multiple Deprivations (IMD; Noble et al., 2004).

A comparison between those women who consented and became part of the extensive sample and those who declined revealed that non-participants were significantly younger ($t(1927) = -5.3$, $p < .001$) and more deprived ($\chi^2(1) = 6.6$, $p < .01$) than those who gave consent.

6.2.2.3 Study phases

Assessments reported as part of this project represent a subgroup of a wider assessment battery used in WCHADS and only elements relative to the current project are described here. In addition to consent and data collection at phase 1 (20-weeks gestation) the extensive sample completed phases of data collection at 9-12 weeks postpartum (phase 5: postal questionnaire) and 14 months postpartum (phase 7: postal questionnaire or face to face).

To align with BCHADS naming conventions, the antenatal assessment will be referred to as T1, the 9-12 week assessment will be referred to as T5, and the 14 month assessment will be referred to as T8. Demographic data was collected in pregnancy (Phase 1/T1) and EPDS questionnaire data at 9-12 weeks (Phase 5/T5) and 12-14 months (Phase 7/T8) postpartum.

6.2.2.4 Follow-up rates from birth to age 14 months

The extensive sample comprised 1,233 mothers with live singleton babies who were eligible for follow-up.

Follow-up rates for the extensive sample were good. In the WCHADS sample a total of 885 (71.2%) mothers completed the EPDS at the 8-week assessment and 828 (67.2%) completed the EPDS at the 12-month analysis and so could be included in analyses. The final samples are compared with the final UKAV sample on key demographic factors in Table 6.2.3 below.

6.2.2.5 Ethical approval

Ethical approval for phases 1 to 7 (T1 to T8) of data collection on the WCHADS was granted by the Cheshire North and West Research Ethics committee on the 27th June 2006 (reference number 05/Q1506/107). The letters confirming ethical agreement for these phases of study are in Appendix 1. Participants gave written informed consent for data collection at multiple phases within the

WCHADS. Information sheets and consent forms that are relevant to the current thesis are given in Appendix 2.

6.2.3 UK Anchoring Vignette (UKAV) Cohort design and sample

The UKAV study was an online study with an observational and cross-sectional design. A convenience sample was recruited to the study between January 2020 and March 2020. In order to match the BCHADS AV subsample and to provide adequate power to perform the AV analysis, the study aimed to recruit 300 participants. This also allowed for some participant error and subsequent removal of data from the sample. Mothers were eligible for inclusion if they were living in the UK, aged 18 years or older, could read or understand spoken English, and had at least one child aged 0-3 years. At the point of invitation (e.g. via online study advert), participants were provided with a brief summary of the proposed study, contact details of the principal investigator for further enquiries about the study, and the link to the web page containing the survey. After accessing the link, the first page showed the participant information sheet and participants were asked to confirm that they had read it. On the next page, participants were asked to provide their informed consent for the study. If they did not consent, the questionnaire automatically skipped to the end of the survey. Participants were recompensed for their time taken to participate.

Study adverts were placed on a number of relevant social media platforms (e.g. Facebook, Reddit) and parenting forums (e.g. Mumsnet, Netmums). 147 mothers were recruited using this approach. However, during regular quality checks the author discovered a number of suspicious responses and further investigation indicated that a number of malicious 'bots' had gained access to the survey. From the point at which the bots first gained access to the questionnaire, it is estimated that approximately 53% of responses being received were malicious. This data was removed from the survey dataset. Following this, the method of data collection was adapted and an ethical amendment was approved to begin collecting data through Prolific.com. This is a research participation website that uses a stringent pre-screening process and has a number of inbuilt features to detect suspicious questionnaire responses. The eligibility criteria were submitted to the service, as well as the number of participants required, and the questionnaire was made available only to those who were eligible. Once the requisite number of participants had completed the questionnaire, the data collection was automatically stopped. 164 participants were recruited to the study using this method.

A total of 311 mothers completed the AV assessment. However, this sample was reduced after removing participants who indicated their child was younger than 6 months. This decision was taken due to differences in reporting behaviour in the early (8 weeks) and later (6-24 months) postnatal

period, made apparent by stark difference in EPDS cut-points established by the Indian validation study described in Chapter 3, Section 3.2.2.1. Following this the sample was further reduced after removing participants who had indicated that their child was older than 36 months, as this was the cut-off established in the study protocol. Thus, the final UKAV sample that was included in the analysis was 226.

Table 6.2.3 shows the key demographic factors for each of the included UK-based samples. Of particular importance is the comparison between the WCHADS T8 cohort and the UKAV cohort as these intended to be as closely matched as possible for the AV analysis. The mean age UKAV mothers was similar but slightly higher than that of the WCHADS T8 mothers. There was also a higher proportion of UKAV mothers who had received at least 5 GCSEs and who were married or cohabiting, while there was a lower proportion of UKAV mothers who were employed and who were white-British.

Table 6.2.3: Maternal characteristics in the final UKAV sample and the final WCHADS T5 and T8 samples.

Characteristic		WCHADS T5	WCHADS T8	UKAV
N		885	828	226
Mother Age	(M, SD)	28.26 (5.84)	29.61 (5.68)	31.7 (5.15)
Employed	(Yes)	85.9%	85.3%	73.0%
At least 5 GCSEs	(Yes)	75.3%	71.5%	84.9%
Ethnicity	White-British	95.7%	96.6%	83.6%
	Other	4.3%	3.4%	16.4%
Partner Status	Married/ Cohabiting	82.2%	81.6%	91.6%
	Single	7.8%	9.1%	7.1%
	Other	10.0%	9.3%	1.9%

6.2.3.1 Ethical approval

Ethical approval for the UKAV study was granted by the University of Liverpool Research Ethics Committee on 13/01/2020 (ref: 5741), and an amendment to the study protocol was approved by the same on 25/03/2020 (see Appendix 1). Copies of the information sheets and consent forms for the study can be found in Appendix 2.

6.2.4 Measures

6.2.4.1 Maternal depression

Maternal depression was assessed in each cohort using the Edinburgh Postnatal Depression Scale (EPDS) (Cox, Holden, & Sagovsky, 1987), which has been described in detail in chapter 3.

6.2.4.2 Differential item functioning (DIF)

DIF was assessed in the BCHADS and UKAV cohorts using a novel set of anchoring vignettes that have been developed as part of this project for use with the EPDS. The vignettes are described in detail in chapter 5.

6.2.5 Procedure

Procedures were conducted differently in each cohort and so will be described separately.

6.2.5.1 BCHADS procedure

At T5 and T8 the EPDS was administered as part of a battery of maternal and child questionnaires. Due to low levels of literacy in the sample, research assistants administered the scale orally. Each item was read out followed by the response options. Participants were asked to select a response which was written down by the researcher.

The anchoring vignettes were administered orally at T8 or T9 and data was entered using the Qualtrics offline app on an android tablet. Participants were each asked to rate 20 vignettes, with 2 corresponding to each EPDS item. Initially, an adaptive approach was used to automatically select vignettes based on the participants' self-rating given for each EPDS item. This was done by ranking the severity of the vignettes within each set *a priori* and then programming Qualtrics to present the two vignettes ranked either side of the self-rating to participants. However, this approach was adapted part way through administration. Low levels of symptom endorsement on the EPDS meant that some responses were rarely or never endorsed and as a result some of the AVs were not being presented with a high enough frequency. Therefore, Qualtrics was reprogrammed at this point to present 2 random vignettes from within each set. Note that the procedure experienced by participants did not change at all throughout administration.

Prior to conducting assessments, researchers were provided with a standardised set of procedures and a troubleshooting guide for common challenges that had emerged during piloting. All researchers took part in group training and conducted practice assessments. Researchers began the assessment by introducing and explaining the task to the participant using a pre-written script. As this was an unfamiliar task, participants completed a practice prior to the actual task. At the

beginning of the assessment, the researcher read out a standardised set of instructions which asked participants to consider the character to be similar to them, the same age and background as them, to imagine themselves in the characters' position, and to pay attention to how the character is feeling and how long they have been feeling that way. This standard set of instructions was then re-read at regular intervals, every three vignettes, to ensure they were followed. Each vignette was presented in a standardised format as follows. Researchers would say, *"We will now listen to the first (next) story and afterwards I will read out 4 options that describe how that person has been feeling in the past 7 days"*. The vignette would then be played and the researcher would read out the corresponding question and response options. The researcher would then say, *"I will now play the vignette again to make sure you have heard all the details and so you can choose the option that best describes the character".* Following the second listening, the researcher would read the response options again and ask the participant to choose a response. Participants were also told they could listen to the vignettes as many times as necessary. The vignettes were presented in a random order.

6.2.5.2 WCHADS procedure

The EPDS was completed at T5 and T8 as part of a set of postnatal questionnaires that were posted to all mothers in the extensive sample, which mothers completed independently and then returned in a freepost envelope. Non-responders were followed up with a phone call and/or a second posting if appropriate.

6.2.5.3 UKAV procedure

The EPDS and the Anchoring Vignettes were administered in the UK online using Qualtrics software (Qualtrics, 2019). The EPDS was completed first. While procedures were kept as similar as possible to the administration in BCHADS, the online format required that some modifications be made. Both the EPDS and AVs were self-administered. This is closer to the recommended format and was more akin to how the EPDS was administered to the WCHADS sample. The UKAV administration used a randomised format to present two random vignettes from each set from the outset. AVs were pre-recorded and uploaded to Qualtrics by a single researcher to ensure that participants were able to listen to oral version of vignettes if required.

A description of the task and instructions for completion were provided in the information sheet at the beginning of the study. Participants were then asked to complete a short demographic questionnaire followed by the EPDS in the standard format. Further instructions were then provided regarding the anchoring vignette task. Instructions mirrored those given to the BCHADS participants. Due to the limitations of the Qualtrics software, it was not possible to repeat the instructions after

every three AVs and maintain the randomised presentation. So instead, the survey was programmed to repeat the instructions after 10 vignettes had been rated. Participants were presented with 20 AVs in total. The vignettes were presented in a random order.

6.3 Analysis Strategy

6.3.1 Measurement invariance testing

Measurement invariance for the Edinburgh Postnatal Depression Scale (EPDS) was tested in two ways. First, to test and compare longitudinal invariance between T5 and T8 within each cohort, two distinct series of longitudinal item factor analyses were performed, one series for BCHADS and one series for WCHADS. Second, to test cross-cultural invariance between the UK and Indian cohorts, a series multi-group item factor analyses were performed at two separate time-points (T5 and T8). Each approach will be described in turn.

6.3.1.1 Longitudinal item factor analyses (LIFA)

In model 1 (configural model) the pattern of factor loadings was constrained to be the same across time-points to test whether all items at each time-point would load onto a single factor (T5 Depression or T8 Depression). In model 2 (metric model) individual factor loadings were constrained to be equal to test whether the contribution of the individual items would vary by phase (weak factorial invariance). In model 3 (scalar model) the individual thresholds were constrained to be equal to examine whether the items performed equivalently across time-points (strong factorial invariance). Models were repeated for each cohort.

LIFA models commonly automatically constrain the factor loading of the first item entered into the model for model identification purposes, in this case EPDS item 1. In order to test that results were not influenced by this constraint, LIFA models were re-run by entering the item that showed the smallest difference into the model first (EPDS item 9) for model identification purposes, and the factor loadings for item 1 were allowed to vary. Results were not substantially different.

6.3.1.2 Multi-group item factor analyses (MIFA).

Model testing followed the same pattern as the LIFA testing but across cohorts instead of time-points. Models were repeated for data at T5 and T8.

6.3.1.3 Approach to invariance testing and model estimation in LIFA and MIFA

Full invariance was demonstrated when the placing of additional constraints on the model did not produce a significant worsening of model fit. The *LRTEST* command was used to evaluate whether a significant change in model fit occurred as a result of imposing additional constraints required for

each model. A non-significant chi-square difference test is considered indicative of invariance (Cheung & Rensvold, 2002). If a significant chi-square value was found, individual items were tested to determine the items with the largest differences between cohorts or time-points. Starting with the item with the largest differences, individual constraints were relaxed one item at a time and the model fit was re-examined as a test of partial invariance. If partial weak invariance was found, strong invariance could only be partial at best. If partial weak invariance was not achieved, then we did not proceed to test strong invariance.

Depression data in community samples is traditionally very skewed, and the current data is no exception. Endorsement of depressive symptoms was low, with all items showing <10% endorsement of non-zero responses. Therefore, to avoid numerical problems associated with low endorsement of responses at the high end of the scale, responses to each item were dichotomised as no (0) vs yes (1, 2, or 3) and treated as ordinal. Following dichotomisation, the latent depression score was treated as normal and models were estimated using maximum likelihood. This approach addresses data missing at random by using available data to compute maximum likelihood estimates for those values. All models were estimated using the generalised structural equation model (GSEM) command in Stata version 16 (MIFA) and 14 (LIFA). General fit indices are not available with the GSEM command in Stata so could not be used to test for the adequacy of GSEM models.

6.3.2 Anchoring vignette analysis strategy

Figure 6.3.2 shows a normal distribution across an underlying latent depression scale divided by thresholds corresponding to the EPDS response scale. The anchoring vignette model allows us to observe the proportion of the distribution that lies between the thresholds. These thresholds represent the implicit cut-points between categories that respondents are using to select options on the response scale. The probability of a given response being selected is a function of the proportion lying between thresholds and this can vary by changing the mean of the distribution (0 in the figure), changing the spread (variance) of the distribution, or by changing the positions of the thresholds, either as a group or individually. In this case the groups are the different countries of origins and may differ in their latent means, the spread (variance) of depression scores, or in the locations of the thresholds, referred to as “country response style”. Individuals can differ in where they “sit” within their country’s distribution but may also differ in where they locate the thresholds, referred to as “individual response style”.

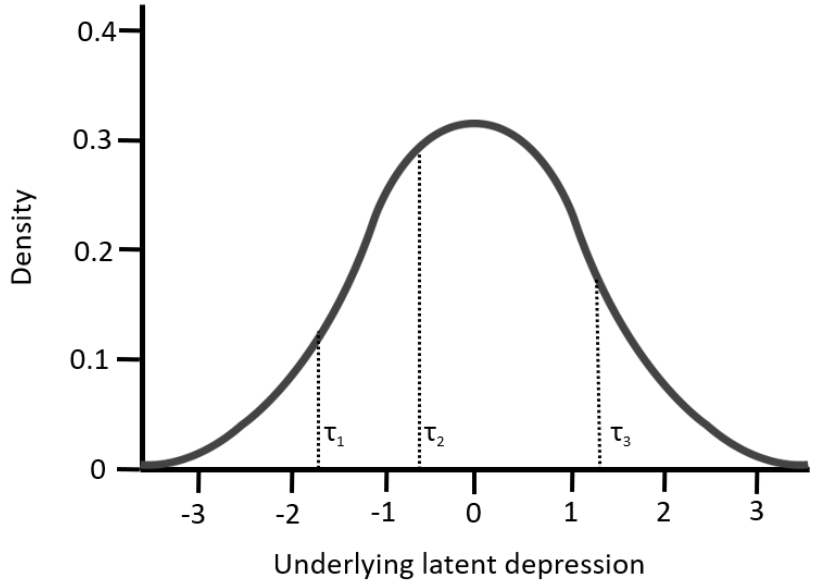
Different models can be applied that allow for some or all of these variations in different ways. Due to the exploratory nature of this study and the novelty of the current approach, we were not in a position to apply a single comprehensive model for which the method of adjustment was known to

be correct. Instead, we aimed to test two distinct models that incorporate different adjustments, and assess the evidence for them, along with their plausibility and impact on comparisons of postnatal depression data between India and the UK.

Briefly, both models allow for variation in response style at a country and individual level, and both allow for differences in factor loadings between items, meaning that different symptoms may have more influence on the latent mean depression score, but constrains them to be the same across countries. Individual-specific bias is modelled in the same way in each model by allowing an individual-specific upward or downward shift in the severity that an individual reports for a given level of depressive symptoms. The models differ in terms of how adjustments are made for country-level bias. In the first model bias is assumed to be shared across items and so thresholds are shifted by a uniform mean bias (difference) between countries that is detected from differences in the locations of vignettes between countries. This is equivalent to a common shift of all the thresholds for each item in one country relative to the other, rather than allowing unique variation in individual thresholds between countries. This model will be referred to as the "mean shift" model. In the second model, bias is detected by fixing the positions of vignettes across countries but allowing thresholds to vary arbitrarily between countries for each item. In both models, the resultant adjustment is applied to the latent depression factor and adjusted factor scores are then extracted for individuals in both countries. This model will be referred to as the "free threshold" model.

The mean shift model is described first and provides the detail of how the AV model is specified. Next the free threshold model is described but as it shares many of the same foundational aspects as the mean shift model, only a brief summary is required to highlight key differences.

Figure 6.3.2: Example of a normal distribution of latent depression scores divided by response thresholds (τ)



6.3.2.1 The Mean Shift model – model specification

The different response sets of the EPDS items and multiple thresholds per item required more AVs than is typical, and the embedding of the study in an ongoing cohort study limited the sample size to be smaller than previous studies (e.g., Kang & Grol-Prokopzcyk, 2020; Knott, Lorgelly, Black, & Hollingsworth, 2017; Molina, 20170). As a consequence, we adopted an approach to the analysis that began with a standard Item Response Theory Model that was as economical as possible on the number of additional parameters included in the model.

The aim was to estimate the underlying depression score (θ) of individual j from country h as a function of covariates describing the individual in a vector X_{hj} and a covariate adjusted mean difference between countries. We assumed these underlying scores follow a conditionally normal distribution with a common variance between countries, and that the effects of any covariates are assumed linear.

$$(1)\ G(\theta_{hj}) = N(\beta X_{hj}, \sigma)$$

The underlying trait θ is considered as linked to an unobserved *continuous* score on each symptom as assessed by item i, Y^*_{hij}. As in a factor model, items can differ in how responsive they are to the underlying depression, and this variation is captured in a factor loading that in this context is referred to as the item discrimination parameter λ. We assumed a common factor loading for each item across countries so that the underlying depression construct is reflected in the items in the same way in both groups. Equation 2 shows that the unobserved continuous score Y* assessed by item i for individual j in country h, is a function of underlying trait θ and a common item discrimination parameter λ across countries.

$$(2)\ Y^*_{hij} = \lambda_i \theta_{hj}$$

The EPDS response sets have four categories (K=4). We assumed an ordinal or graded response model that maps the unobserved continuous score for each item to the observed response category Y_{hij}=k. The probability of the unobserved continuous score mapping to observed category k depends upon the proportions of the normal curve lying between the K-1 thresholds of the item and is therefore influenced by the location of those thresholds or the mean of the distribution of responses in each group.

$$(3)\ Pr(Y_{hij} = k) = \Phi_k(Y^*_{hij} > \tau_{hik}) - \Phi_{k-1}(Y^*_{hij} > \tau_{hi(k-1)})$$

where Φ is the normal cumulative distribution function. The mapping of the unobserved continuous score to the category reported by the respondent may vary as different individuals may interpret the categories as requiring more or less of the underlying score to meet the requirements of the response set threshold. These differences we characterise as *bias* and it is in how this bias is conceptualised that the two AV models diverge. In the mean shift model, bias is assumed conditionally independent of θ, the underlying depression, and may be a function of observed covariates X_{hj} and additional individual differences attributed to a latent bias factor ν whose impact may vary from item to item but is assumed to be Gaussian. Thus

$$(4)\tau_{hijk} = \tau_{ik} + \text{bias}_{hij}$$

$$(5)\tau_{hijk} = \tau_{ik} + \delta_{hi}X_{hj} + \gamma_i v_{hj}$$

$$(6)\ v_{hj} = N(0, \sigma)$$

In the absence of data from anchoring vignettes the parameters to separately identify levels of depression and bias cannot be jointly estimated.

6.3.2.2 Mean Shift model - model estimation

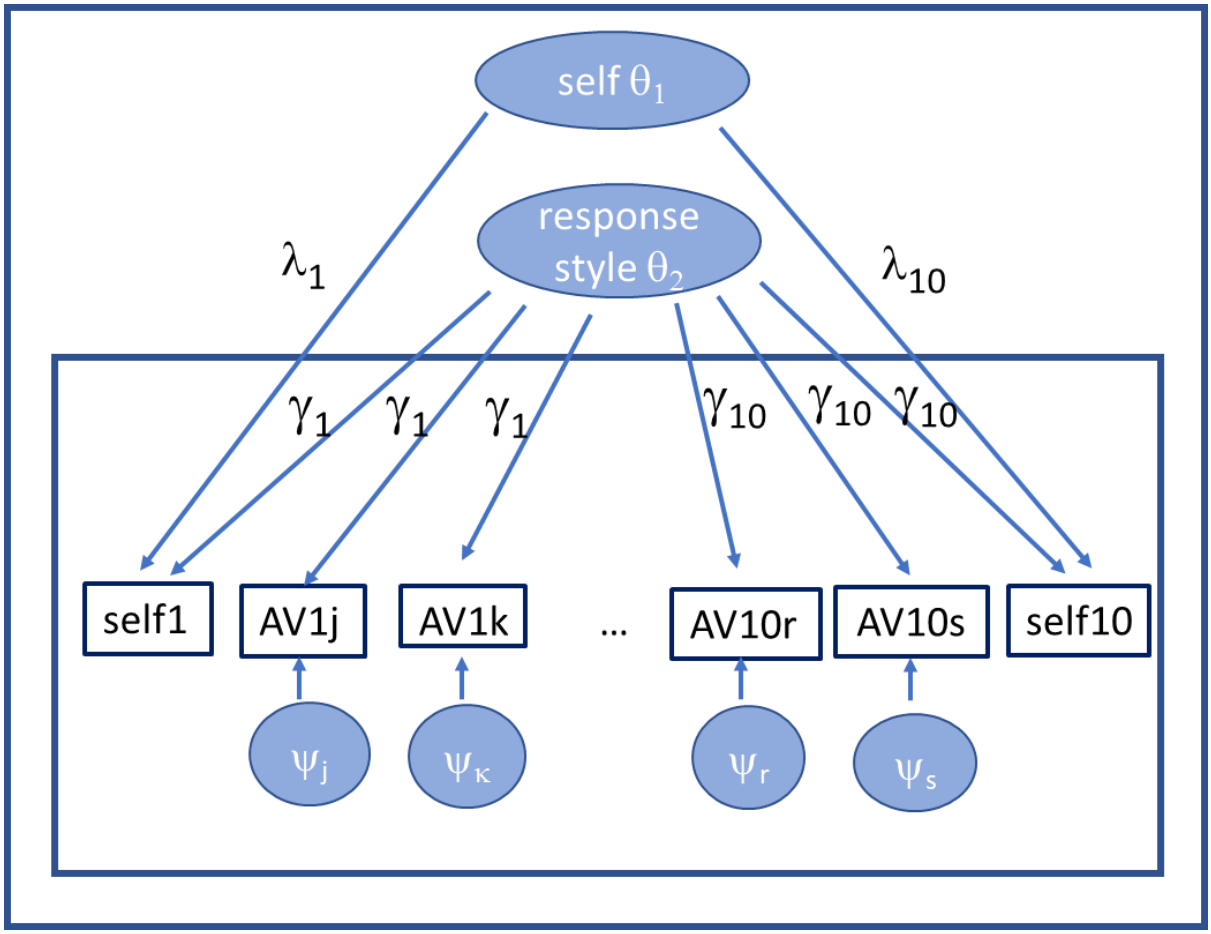
Standard IRT model estimation software is not able to analyze both individual and AV data. The probit based Graded Response Model treats the subject specific underlying depression score as an unknown latent variable (or generalized random effect). When probit based IRT models are estimated by maximum likelihood an integrated likelihood is computed that calculates the likelihood of the observed response by numerical quadrature, an average of the likelihoods of the observed response for different possible values of the underlying depression score, weighted by their probability of occurrence.

While for each item we have just a single subject report, each AV received multiple ratings by many individuals drawn from both countries. For these, a fixed effects approach was adopted, with a parameter corresponding to the position of the vignette on the underlying scale being estimated for each of the AV's separately for each country. The difference in the positioning of the same AV's by country and individual is then taken as evidence of country-level bias and applied to the latent depression mean to correct for DIF.

Responses to both self-report and anchoring vignettes are a function of these fixed effects and the latent factor for subject specific bias. The self-reports are additionally a function of the latent depression factor. To ensure that the thresholds for both anchoring vignettes and self-report divide the cumulative distribution of the underlying Y* in the same way, the variances of the underlying

distributions for AVs and self-reports must be made equal. This is achieved by the addition to the term contributing to the AV responses of a set of vignette-specific errors with the same variance as that of the variance contribution from the depression factor for the self-reports for that item (see Figure 6.3.2.2).

Figure 6.3.2.2: Multilevel structural equation model for AVs: illustrating how response thresholds for AVs and self-reports are estimated after constraining the self-reports and vignettes to share the same variance.



Note - θ_1 = latent depression factor; θ_2 = subject specific bias factor; Ψ = vignette-specific error with same variance as latent depression factor; self1…10 = EPDS items 1-10; AV1…10 = AVs corresponding to each EPDS item.

A gsem model that fits this model to the BCHADS (India) and WCHADS (UK) sample data with country specific effects on depression score and bias but with common loadings across countries for both depression and bias factors is shown below. There are 10 items and the dummy variable *meas* distinguishes which of the self-report or anchoring vignettes (there are 6 possible vignettes per item) the response pertains to. Using the default `gsem` notation with capitals for latent variables, the depression factor G loads only on self-report (with factor loadings a-j), while the subject-specific bias factor F loads on all responses with loadings k1-k10.

```
gsem  (av_1_<- i.meas uk c.g#G[newid]@a c.gnot#H[newid>meas]@a F[newid]@k1 , oprobit ) ///
      (av_2_<- i.meas uk c.g#G[newid]@b c.gnot#H[newid>meas]@b F[newid]@k2 , oprobit ) ///
      (av_3_<- i.meas uk c.g#G[newid]@c c.gnot#H[newid>meas]@c F[newid]@k3, oprobit ) ///
      (av_4_<- i.meas uk c.g#G[newid]@d c.gnot#H[newid>meas]@d F[newid]@k4 , oprobit ) ///
      (av_5_<- i.meas uk c.g#G[newid]@e c.gnot#H[newid>meas]@e F[newid]@k5, oprobit ) ///
      (av_6_<- i.meas uk c.g#G[newid]@f c.gnot#H[newid>meas]@f F[newid]@k6 , oprobit ) ///
      (av_7_<- i.meas uk c.g#G[newid]@g c.gnot#H[newid>meas]@g F[newid]@k7, oprobit ) ///
      (av_8_<- i.meas uk c.g#G[newid]@h c.gnot#H[newid>meas]@h F[newid]@k8 , oprobit ) ///
      (av_9_<- i.meas uk c.g#G[newid]@i c.gnot#H[newid>meas]@i F[newid]@k9, oprobit ) ///
      (av_10_<- i.meas uk c.g#G[newid]@j c.gnot#H[newid>meas]@j F[newid]@k10 , oprobit ) ///
      (G[newid] <-uk) , iter(50) var(e.G[newid]@1) var(H[newid>meas]@1) var(F[newid]@1)
```

6.3.2.3 Free-Threshold model – brief specification and estimation

This model begins from the perspective that, provided that the scale of Y* is the same across both countries, then, assuming vignette equivalence, the fixed positions of the anchoring vignettes are also shared between countries. The thresholds for each item $\{ \tau_{hi(k)} \}$ are then estimated separately for each country as a function of the variation in actual ratings given to the AVs between countries. We additionally allow for variance between individuals in the use of thresholds by the inclusion of the same subject specific bias factor as in the previous model. The mean of the self-rated depression is located entirely independently in each country as a function of the self-assessments and the locations of the thresholds in each country.

To ensure that the underlying depression scale Y* is the same for both countries, we constrained the variance of the self-rated depression to be the same across countries and constrained the factor loadings for the latent depression factor to be the same for each item between countries. An additional latent variable was introduced with the same factor loading and variance as G to ensure that the self-ratings and vignette ratings also share the same underlying scale.

With all the thresholds free across both countries, for identification we restricted the mean of the latent depression variable in one country (India), here we chose 0, and measured the position of the vignettes and mean depression in the other country relative to this fixed point. This means that AVs and the mean self-rated depression of the UK sample are estimated relative to this point Consequently, all estimates in the model can be understood in terms of standard deviations from the constrained Indian mean latent depression.

This model can be setup in a multiple group `gsem` model by specifying the country as a group but without the default assumption of the invariance of thresholds.

```
gsem  (av_1_<- i.meas  c.g#G[newid]@a c.gnot#H[newid>meas]@a , oprobit ) ///
      (av_2_<- i.meas  c.g#G[newid]@b c.gnot#H[newid>meas]@b , oprobit ) ///
      (av_3_<- i.meas  c.g#G[newid]@c c.gnot#H[newid>meas]@c, oprobit ) ///
      (av_4_<- i.meas  c.g#G[newid]@d c.gnot#H[newid>meas]@d , oprobit) ///
      (av_5_<- i.meas  c.g#G[newid]@e c.gnot#H[newid>meas]@e, oprobit ) ///
      (av_6_<- i.meas  c.g#G[newid]@f c.gnot#H[newid>meas]@f, oprobit ) ///
      (av_7_<- i.meas  c.g#G[newid]@g c.gnot#H[newid>meas]@g , oprobit ) ///
      (av_8_<- i.meas   c.g#G[newid]@h c.gnot#H[newid>meas]@h, oprobit ) ///
      (av_9_<- i.meas  c.g#G[newid]@i c.gnot#H[newid>meas]@i, oprobit ) ///
      (av_10_<- i.meas  c.g#G[newid]@j c.gnot#H[newid>meas]@j , oprobit ), ///
      group(uk2) ginvariant(loading coef)  var(1:G[newid]@1) var(2:G[newid]@1) mean ///
      (1:G[newid]@0) mean(2:G[newid]@m, init(0)) var(1:H[newid>meas]@1) ///
      var(2:H[newid>meas]@1) mean(1:H[newid>meas]@0) mean (2:H[newid>meas]@0)
```

The `gsem` model code was provided by a supervisor (AP).

6.3.2.4 The Naïve Model

Both models share a common "naïve" restricted version in which it is assumed that the thresholds and factor loadings are invariant and shared between all individuals across both countries. This model provides an unadjusted estimate of underlying latent depression scores, facilitating an unadjusted comparison of mean scores between countries.

6.3.2.5 Treatment of missing data for all models

IRT model estimation by maximum likelihood offered the scope to exploit the property of ignorability. This allows any data whose probability of missingness is dependent only upon data that has been observed, or is otherwise entirely random, to be ignored. This allows testing to be adaptive, where the items or vignettes presented to respondents can vary according to the responses that the respondent has already given. The model also allows multiple vignettes for each item and does not require the anchoring vignettes and self-reports to be obtained on the same sample.

6.3.3 Exploratory analyses to test accuracy of AV models

Further analyses were performed to explore and compare the accuracy of each AV model to the raw and unadjusted ratings given to the AVs. Histograms and mean ratings were produced for the vignettes, first for each set and then for each vignette individually.

Country level differences in anchoring vignette ratings were then investigated, with analyses grouped by EPDS item. The data was converted to a long format with each line containing one response to a single vignette. The vignette that each response applied to was identified by the variable "vignette" which was coded with values corresponding to vignettes A, B, C, D, E and F (1-6). Vignette A represents the *most severe* vignette and vignette F represents the *least severe* vignette. As the data included multiple observations of the same participants, the *xtset* command was used to ensure that it was treated as panel data. The *xtologit* command was then used to test differences in a series of random effects ordered logistic models. Due to the number of tests carried out, we applied a Bonferroni correction to all results ($p<.005$).

6.3.3.1 Main effects of country on vignette ratings.

Model 1a (xtologit rating i.vignette i.country_age if item==1…10): The first test to be carried out examined whether AV ratings for each item differed significantly by country. The dependent variable was coded 0-3, with 0 being the least severe rating and 3 being the most severe. Predictors included a set of dummy variables for the severity of the vignette coded A-F, with 'A' containing a description of the most severe symptoms and 'F' containing a description of the least severe symptoms. "Country" was coded as a binary variable, with 0 = India and 1 = UK. Finally, maternal age was included as a continuous variable.

Vignette coefficients represent severity of rating given to each vignette (B-F) relative to the rating given to vignette 'A'. A significant coefficient indicates that a given vignette is rated significantly differently from vignette A, with a negative sign indicating that the vignette was rated less severely, and a positive sign indicating that the vignette was rated more severely. The 'country' coefficient represents the mean difference in ratings across the vignettes between countries. As India is the reference group, the coefficient specifically refers to whether UK participants rated the vignettes significantly more or less severely than the Indian participants.

6.3.3.2 Moderated effects of country on vignette ratings by vignette severity

Two interaction models were also tested, one including the ordinal variable 'vignette' (model 2), and another including the continuous variable 'severity' (model 3). 'Severity' was generated from

'vignette' and coded from 0-5, with 0 representing the most severe vignette and 5 representing the least severe vignette.

Model 1b (xtologit rating i.vignette i.country i.vignette#i.country if item==1…10).

Model 1b was identical to model 1a but with the additional 'vignette X country' interaction term. A separate analysis was performed for each EPDS item and interaction coefficients are given for each vignette within a given item. The interaction terms can be interpreted as showing whether vignettes shared the same bias, individually and as a group. A significant individual interaction term would suggest that the effect of a given vignette on the rating provided by a participant was significantly affected by the respondents' country group. However, as the vignette coefficients represent the rating for a given vignette compared to that for vignette 'A', the interaction coefficient is likewise relative. Thus, a significant interaction suggests that country level bias is significantly different for a given vignette than it is for vignette 'A'. A non-significant interaction suggests there is no difference relative to the effect of country on ratings of vignette 'A'. In addition, the *contrast* command was used to compare the block of interactions for each item.

Model 2: (xtologit rating i.country c.severity c.mums_age country#c.severity if item==1…10).

Model 2 contained the same dependent variable as model 1 but replaced the set of dummy variables for vignette severity with a single continuous variable, 'severity', and included the 'severity X country' interaction term. A separate analysis was performed for each EPDS item. The interaction term in this case can be interpreted as a more powerful, one degree of freedom, test of a trend in bias. In this model, the main effect of country represents bias at severity = 0 (most severe). The interaction effect represents how bias changes with severity. Note, an increase in the severity variable represents a decrease in the severity of the vignette. Thus, if the main effect of country is positive, then a positive interaction implies that bias increases further as with severity for the UK respondents. Conversely, a negative interaction would imply declining bias with severity increases, possibly passing through no bias and becoming biased in the opposite direction.

All analyses were carried performed using Stata version 14 (StataCorp, 2015) or version 16 (StataCorp, 2019).

6.4 Results

Results from the measurement invariance testing are presented first, followed by results from the anchoring vignette analyses.

6.4.1 Aim 1: Measurement invariance testing results

6.4.1.1 Longitudinal invariance in BCHADS

The 10 items making up the EPDS were tested for measurement invariance between time-points within the BCHADS cohort at 8 weeks (n = 545) and 12 months (n = 549) using longitudinal item factor analysis (LIFA). Model 1, the configural model, showed the same pattern of factor loadings for each item at both time-points. After fixing the factor loadings, model 2a, the metric model, showed a significantly worsened fit (LR $\chi^2(8) = 29.12$, $p < .001$). Individual items were inspected and items 2 and 7 showed significantly different factor loadings between time-points. Theoretical rationale suggested that item 2 *("I have looked forward with enjoyment to things"*) might more reasonably be expected to vary across time-points and so the factor loadings for this item were freed first. With this item allowed to vary, the fit of model 2b was not significantly worse than model 1 (LR $\chi^2(7) = 12.48$, $p = .09$), therefore we found evidence of partial metric invariance. Individual items were then inspected for threshold differences. All items exhibited significant differences and so we did not test for scalar invariance. In the models described above, the factor loading for item 1 at each time-point was automatically constrained for model identification purposes. In order to check that this had not affected results, we identified the item with most similarity in factor loadings between time-points and ran the models again with this item constrained, thereby freeing item 1. Results from this series of models were not significantly different from those reported (see Appendix 9).

6.4.1.2 Longitudinal Invariance in WCHADS

The 10 items making up the EPDS were tested for measurement invariance between time-points within the WCHADS cohort at 8 weeks (n = 885) and 12 months (n = 824) using longitudinal item factor analysis (LIFA). Model 1, the configural model, showed the same pattern of factor loadings for each item at both time-points. After allowing the factor loadings to vary for model 2a, the metric model, model fit was not significantly worse (LR $\chi^2(8) = 14.34$, $p = .07$), therefore we found evidence of metric invariance. After introducing scalar invariance into model 3a, the scalar model, the model fit worsened significantly, as shown by the likelihood ratio test of difference (LR $\chi^2(10) = 69.58$, $p<.001$). Individual items were then inspected to test for variation in thresholds. Items 4, 5, 6, 7 and 9 showed significant differences in thresholds between time-points. In order of the magnitude of difference, items 5, 9, 7 and 4 were freed one at a time to examine whether this improved model fit. After these four items were allowed to vary, model 3e did not show a significantly worsened fit as compared to model 2 (LR $\chi^2(6) = 7.92$, $p = .24$), therefore we found evidence of partial scalar invariance. As previously, factor loadings for item 1 were automatically constrained for model identification. For WCHADS, item 10 showed the least difference between time-points and so we re-

ran the models after constraining this item and results were not significantly different from those already reported (See Appendix 9).

Overall, we found evidence of configural and partial metric longitudinal invariance for 9/10 items, for the EPDS in the BCHADS cohort from 8 weeks to 12 months, but no evidence of scalar invariance. We also found evidence of configural, metric and partial scalar invariance for 6/10 items, for the EPDS from 8 weeks to 12 months in the WCHADS cohort.

6.4.1.3 Age 8 weeks multi-group cross-cultural measurement invariance

The 10 items making up the EPDS were tested for cross-cultural measurement invariance across India (n = 545) and the UK (n = 885) by performing multi-group Item Factor Analysis (MIFA) on the data obtained from the 8-week assessment using GSEM. Model 1, the configural model, showed the same pattern of factor loadings for each item in both cohorts, indicating that the one-factor solution for the EPDS was acceptable for both cohorts. The introduction of factor invariance, Model 2a, the metric model, resulted in a worse model fit as shown by a significant likelihood ratio test of difference (LR $\chi^2(10) = 42.64$, $p<.001$). Individual item factor loadings were then inspected for differences between cohorts. Items 3, 4, 6, 8 and 9 showed significant differences and so these factor loadings were freed. However, even after freeing these items, Model 2b showed a significantly worse fit than Model 1 (LR $\chi^2(5) = 12.93$, $p = .02$). Further tests of difference for individual items showed that item 10 was marginally different ($p =.09$) and so the factor loadings for this item were freed. The subsequent model, model 2c, did not show a significantly worse fit than model 1 (LR $\chi^2(4) = 9.43$, $p =.051$). However, given that only 4/10 items remained unconstrained, we did not meet the threshold for partial metric invariance and so did not proceed to test scalar invariance.

6.4.1.4 Age 12 months multi-group cross-cultural measurement invariance

The 10 items making up the EPDS were tested for measurement invariance across India (n = 549) and UK (n = 825) using MIFA of the data obtained at the 12-month assessment. Model 1, the configural model, showed the same pattern of factor loadings for each item in both cohorts, indicating that the one-factor solution for the EPDS was acceptable for both cohorts. After introducing factor invariance, model 2a, the metric model, showed a worse fit than model 1, with a significant chi-square test of difference (LR $\chi^2(10) = 116.07$, $p<.001$). Individual items were then inspected for significant differences. Items 1, 2, 6 and 8 all showed significantly different factor loadings between cohorts and so were freed for model 2b. With these items allowed to vary, model 2b did not show a significantly worse fit than model 1 (LR $\chi^2(6) = 5.45$, $p = .48$), therefore we found evidence of partial

metric invariance. Inspection of the items at the scalar level revealed significant differences for 7/10 items. Therefore, we did not test the scalar invariance model.

Overall, we found evidence of configural invariance for the EPDS between India and the UK at 8 weeks, but only 4 out of 10 items were invariant at the metric level and so even partial metric invariance was not achieved. We also found evidence of configural and partial metric invariance in 6 out of 10 items for the EPDS between India and the UK at 12 months. There was no evidence of scalar invariance at either time-point. These findings suggest that in order to make direct comparisons of latent means between countries at 12 months, further information is required regarding the item response thresholds used in each sample in order to adjust responses to a common scale. Anchoring vignettes were therefore used to detect differential item functioning (DIF) in the use of response options.

6.4.2 Aim 2: Detection and minimisation of bias using Anchoring Vignettes:

6.4.2.1 Descriptive results for AV data from BCHADS and UKAV.

Visual examination of the histograms showing the density distribution of UK and Indian participant ratings for the individual vignettes for each EPDS item in Figure 6.4.2.1a indicates that the vignettes were generally rated in the expected order and that participants understood the vignette rating task fairly well. A clear distinction between the most and least severe vignettes can be seen in each set, with each corresponding histogram showing a respective right-skewed or left-skewed distribution. The intermediate vignettes generally showed a gradual shift from right to left, indicating that participants understood them as decreasing in severity. The set that appeared to be most problematic related to EPDS item 2, where response patterns indicate that participants may have struggled to distinguish between the severity of vignettes 1 and 2, and between vignettes 3, 4 and 5.

Combined mean ratings for each AV across countries are given in Table 6.4.2.1 and suggest that participants typically perceived and rated the severity of each vignette in the expected order for each EPDS item, with a few minor exceptions. There is one order violation in each set corresponding to EPDS items 3, 5, 7, 8 and 9 but all are quite small and are constrained to the more severe vignettes.

Figure 6.4.2.1a: Histograms showing density distribution of vignette ratings for each EPDS item, split by country.

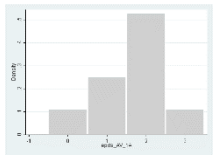
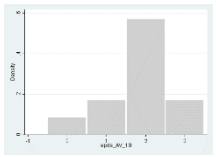
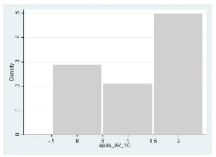
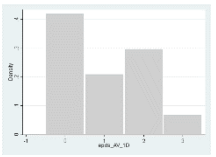
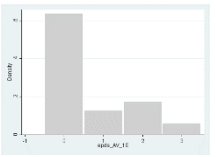
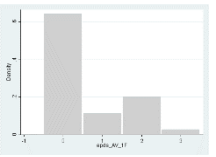
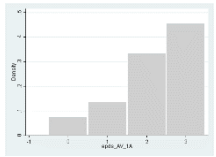
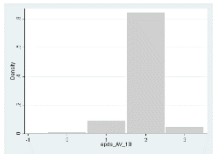
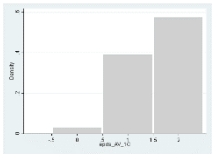
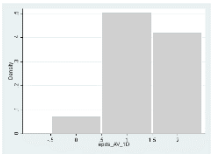
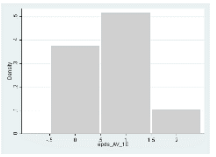
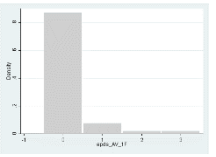
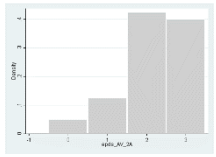
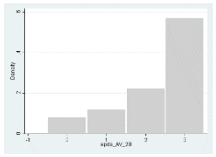
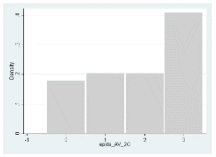
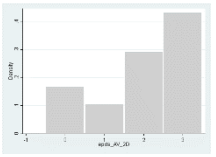
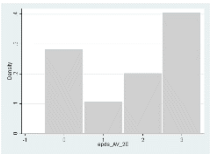
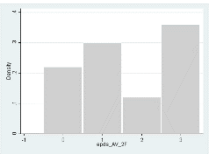
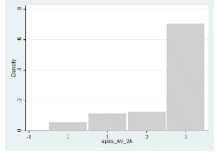
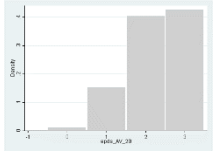
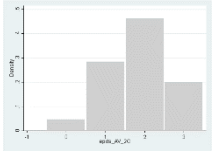
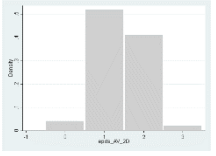
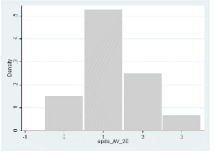
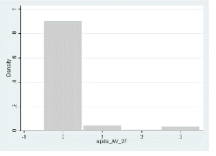
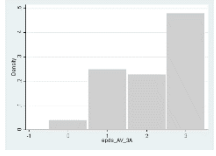
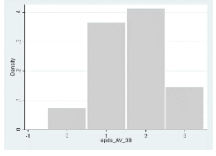
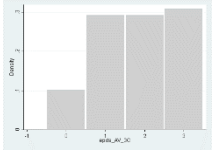
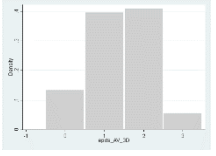
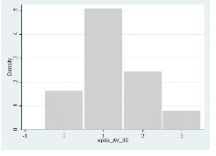
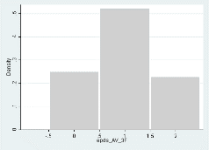
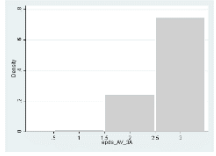
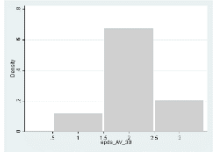
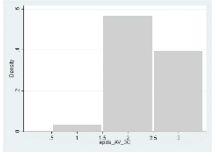
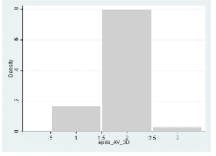
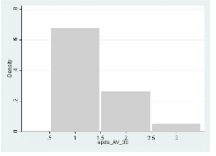
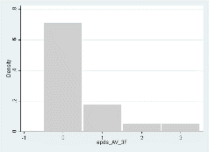
EPDS Item	Cohort	AV-A – Most Severe	AV-B	AV-C	AV-D	AV-E	AV-F – Least Severe
1	BCHADS						
	UKAV						
2	BCHADS						
	UKAV						
3	BCHADS						
	UKAV						

Figure 6.4.2.1a continued: Histograms showing density distribution of vignette ratings for each EPDS item, split by country.

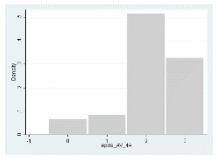
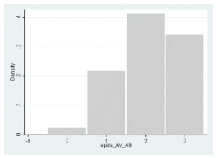
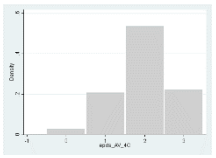
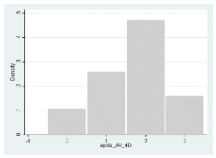
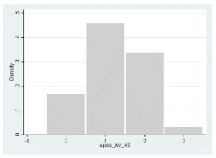
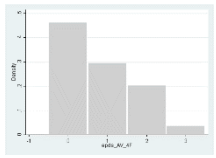
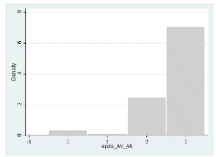
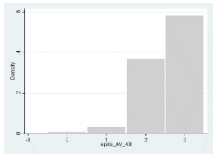
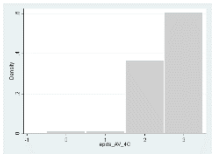
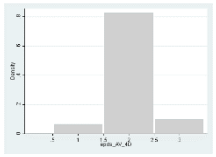
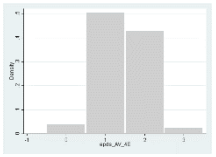
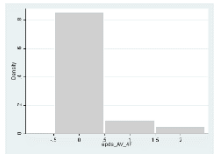
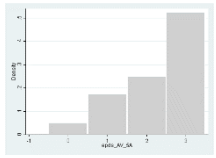
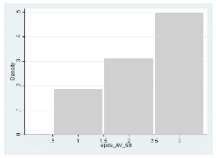
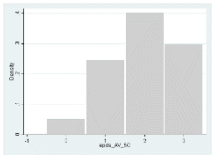
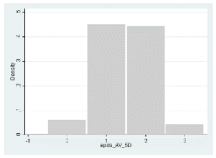
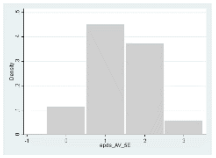
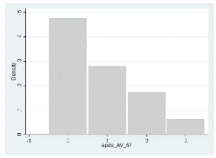
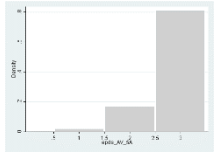
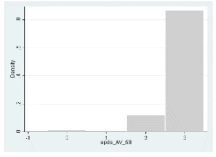
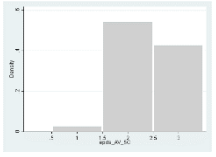
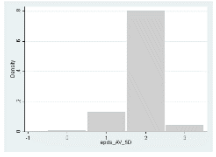
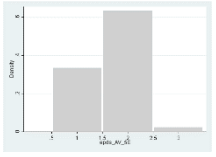
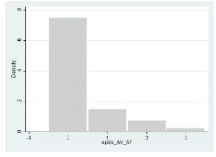
EPDS Item	Cohort	AV-A – Most Severe	AV-B	AV-C	AV-D	AV-E	AV-F – Least Severe
4	BCHADS						
	UKAV						
5	BCHADS						
	UKAV						

Figure 6.4.2.1a continued: Histograms showing density distribution of vignette ratings for each EPDS item, split by country.

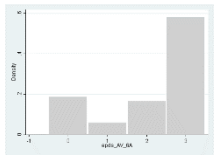
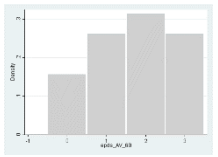
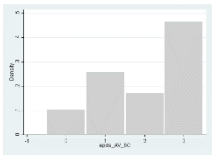
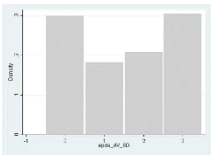
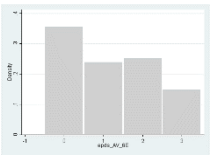
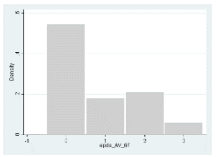
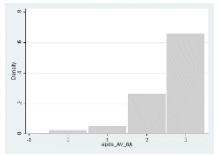
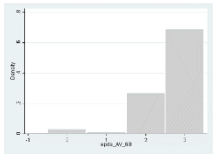
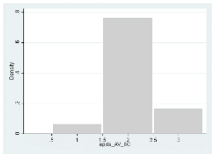
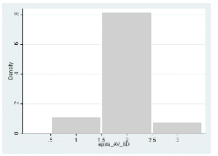
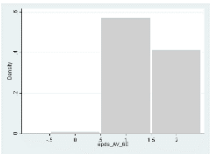
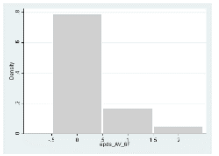
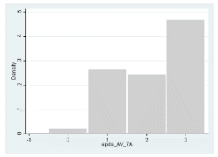
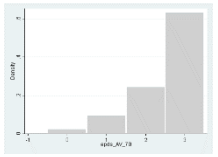
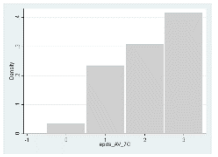
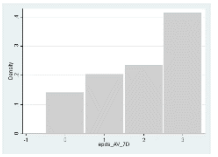
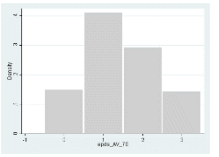
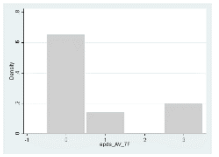
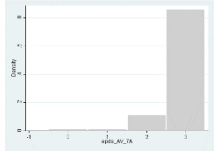
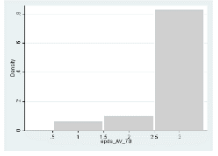
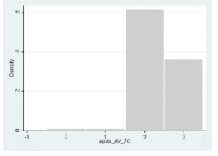
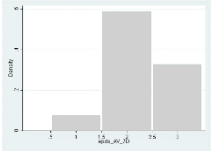
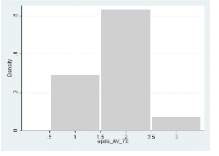
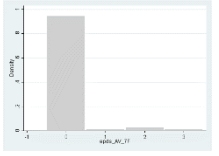
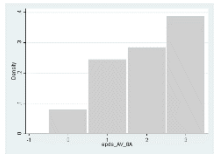
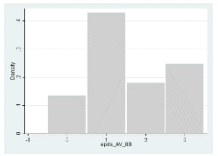
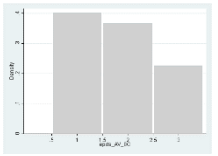
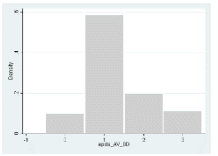
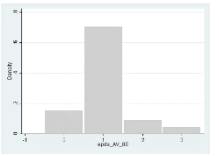
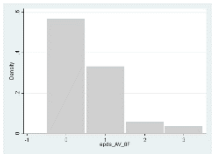
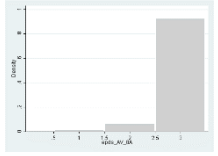
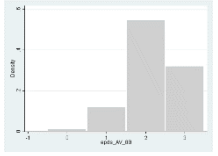
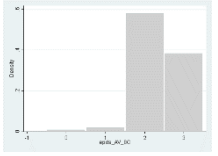
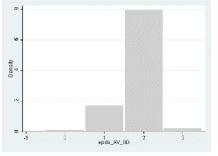
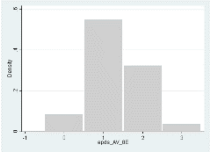
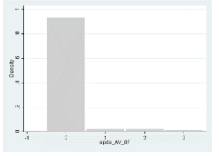
EPDS Item	Cohort	AV-A – Most Severe	AV-B	AV-C	AV-D	AV-E	AV-F – Least Severe
6	BCHADS						
	UKAV						
7	BCHADS						
	UKAV						
8	BCHADS						
	UKAV						

Figure 6.4.2.1a continued: Histograms showing density distribution of vignette ratings for each EPDS item, split by country.

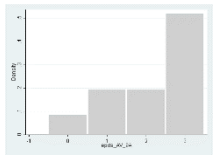
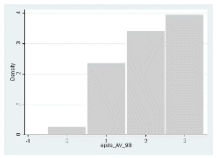
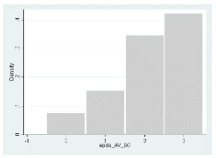
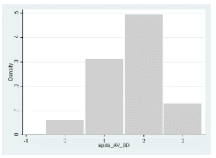
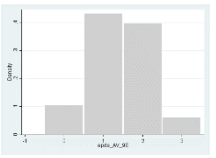
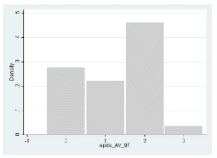
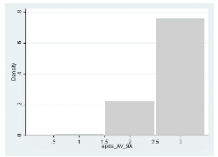
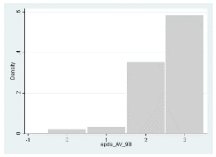
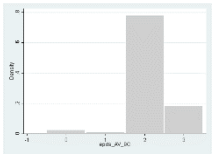
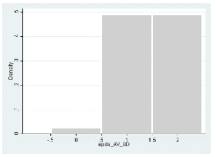
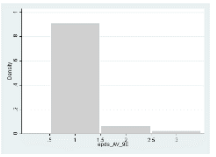
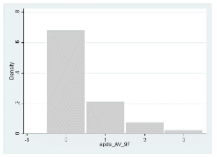
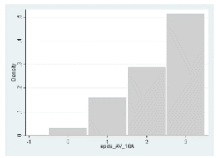
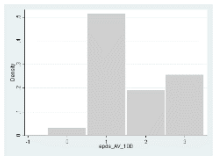
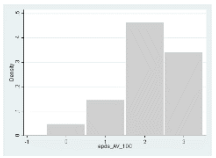
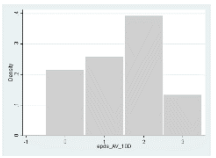
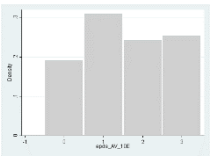
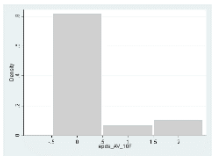
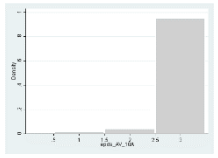
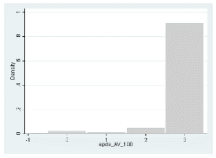
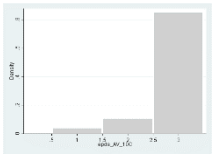
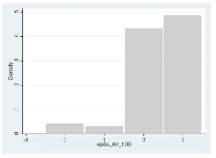
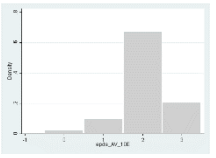
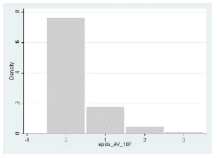
EPDS Item	Cohort	AV-A – Most Severe	AV-B	AV-C	AV-D	AV-E	AV-F – Least Severe
9	BCHADS						
	UKAV						
10	BCHADS						
	UKAV						

Table 6.4.2.1: Combined and separate country mean ratings for AVs corresponding to each EPDS item

Vignette	Item 1		Item 2		Item 3		Item 4		Item 5		Item 6		Item 7		Item 8		Item 9		Item 10	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
A[a]	1.98	(0.93)	2.38	(0.90)	2.52	(0.74)	2.41	(0.79)	2.64	(0.67)	2.40	(0.94)	2.57	(0.75)	2.56	(0.80)	2.55	(0.75)	2.76	(0.59)
B	1.90	(0.57)	2.26	(0.84)	1.95	(0.69)	2.38	(0.72)	2.70	(0.61)	2.34	(0.90)	2.67	(0.65)	1.96	(0.86)	2.38	(0.76)	2.52	(0.85)
C	1.45	(0.68)	1.83	(0.93)	2.14	(0.81)	2.29	(0.73)	2.20	(0.75)	2.06	(0.79)	2.25	(0.71)	2.15	(0.70)	2.12	(0.72)	2.56	(0.71)
D	1.13	(0.90)	1.79	(1.00)	1.57	(0.71)	1.82	(0.75)	1.62	(0.64)	1.67	(1.04)	2.05	(0.95)	1.50	(0.74)	1.61	(0.71)	1.75	(1.01)
E	0.68	(0.88)	1.56	(1.14)	1.29	(0.75)	1.31	(0.72)	1.48	(0.71)	1.27	(0.92)	1.55	(0.83)	1.13	(0.68)	1.31	(0.67)	1.73	(0.97)
F[b]	0.31	(0.71)	0.72	(1.12)	0.65	(0.82)	0.43	(0.74)	0.54	(0.85)	0.41	(0.73)	0.31	(0.84)	0.30	(0.66)	0.77	(0.91)	0.30	(0.63)
Mean																				
Combined	1.24		1.76		1.69		1.77		1.86		1.69		1.90		1.60		1.79		1.94	
India	0.99		1.91		1.45		1.60		1.55		1.50		1.78		1.31		1.69		1.54	
UK	1.29		1.58		1.85		1.91		2.03		1.81		2.04		1.82		1.78		2.20	
Difference	.30		-.33		.40		.31		.48		.31		.26		.51		.09		.56	
Range																				
Combined	1.67		1.66		1.87		1.98		2.16		1.99		2.36		2.26		1.78		2.46	
India	1.21		0.67		1.17		1.29		1.49		1.36		1.74		1.41		0.89		2.00	
UK	1.98		2.30		2.28		2.43		2.47		2.35		2.71		2.79		2.31		2.63	
Difference	.77		1.63		1.11		1.14		.98		.99		.97		1.38		1.42		.63	

Note – [a] most severe AV; [b] least severe AV.

Mean AV ratings are also presented separately for each country and for each EPDS item in Figure 6.4.2.1b to allow a more detailed visual examination of the trends in rank ordering of AVs within each country. Similar to the combined mean ratings in Table 6.4.2.1, these charts indicate that ordering violations were constrained to the more severe vignettes (vignettes A to C). Mean ratings for vignettes D, E and F indicate that these vignettes were always rated in the expected order, both when countries were considered together and separately. They also indicate that UK participants rated the vignettes in the expected order more consistently than Indian mothers. However, while not perfectly ordered, Indian mothers did generally rate the vignettes as expected and there was clear trend of decreasing ratings with decreasing severity.

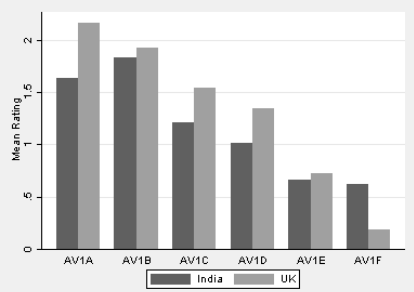
There were also a number of patterns observable in the data that suggest Indian and UK respondents mis-ordered vignettes in the same way. For example, while there was some mis-ordering of vignette severity evident for EPDS items 3, 8 and 4, the patterns are similar across countries, indicating that the vignettes were being perceived in the same way.

However, there were also some inconsistent ordering patterns between countries, as seen via a visual inspection of items 1, 6 and 7. These violations are relatively minor however, and only items 2 and 10 show consistent mis-ordering of vignettes.

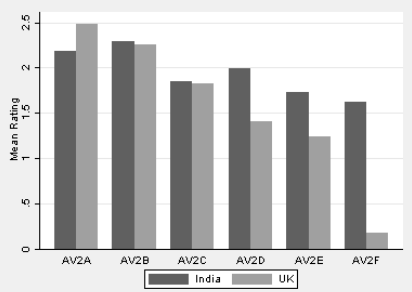
Overall, the mean ratings indicate that vignettes were generally perceived to represent equivalent levels of underlying depression (even if different thresholds were applied to them). Other than items 2 and 10 there was a consistent trend in the mean ratings of the vignettes which showed that participants in both groups perceived them as decreasing in severity in the expected way. There were more violations in the more severe vignettes and in the Indian sample, but in most cases these violations were small, indicating that participants may have had difficulty in distinguishing between vignette severity when rating the more severe vignettes, rather than perceiving them in the incorrect order.

Figure 6.4.2.1b: Mean AV ratings (0-3) for India (dark grey) and UK (light grey) separately for each EPDS item set of vignettes.

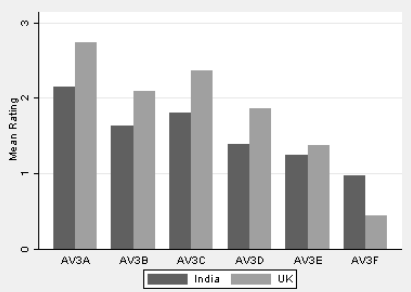
EPDS Item 1



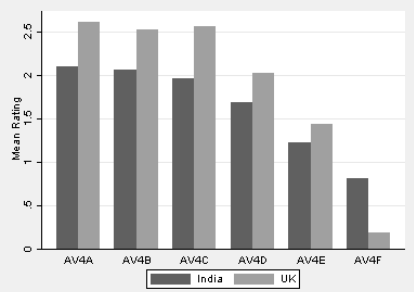
EPDS Item 2



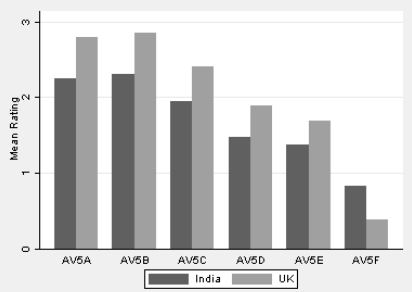
EPDS Item 3



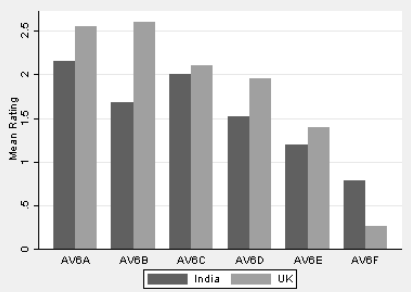
EPDS Item 4



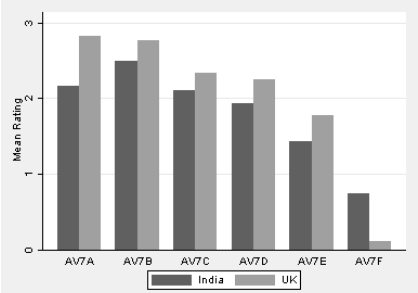
EPDS Item 5



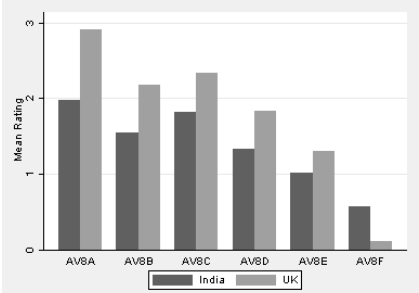
EPDS Item 6



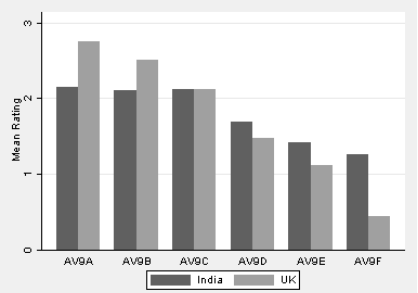
EPDS Item 7



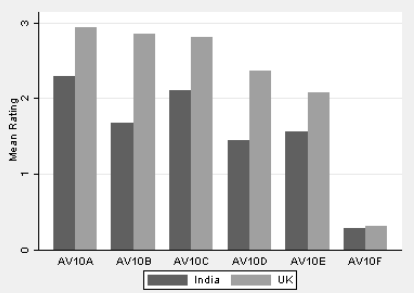
EPDS Item 8



EPDS Item 9



EPDS Item 10



6.4.2.2 Results from parametric analysis of AV data from BCHADS, UKAV and WCHADS.

Results of the main anchoring vignette analyses are now presented and estimates are compared between three models. The first is the "naïve" model in which latent depression scores were estimated from self-assessments alone and which provides a comparison between India and the UK that has not been corrected for differential item functioning (DIF). The second model is the "mean shift" model in which latent depression means were estimated from self-assessments and AV ratings. The third model is the "free threshold" model in which DIF was estimated separately in each item by allowing thresholds to vary between countries uniquely for each set of AVs as a function of variation in AV ratings. Results from the naïve model are presented first, followed by results from the mean shift and free threshold models. The results from the two AV adjusted models are presented in parallel to facilitate comparison. Following this, the strengths of each model of are discussed in the context of additional exploratory analyses.

6.4.2.3 Naïve model

When estimated from self-assessments alone, there was a significant and medium-sized country difference in latent depression means (d = .46, $p < .001$), with UK mothers reporting higher rates of postnatal depression than Indian mothers at 12 months. Model estimates are presented in Table 6.4.2.3 and show that each item made a significant contribution to the underlying depression score. Cut-points varied between items, indicating that participants were applying unique thresholds to the response scale for each item.

Table 6.4.2.3: Summary statistics from naive model, including the factor loading of each item onto the Latent Depression Factor, the response thresholds for each item, and the unadjusted mean difference in Latent Depression between countries.

Shared Coefficients	**EPDS Item**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Latent Depression Factor	1[a]	.298**	.982**	1.141**	1.206**	.991**	1.596**	2.124**	1.579**	1.068**
Cut 1	1.896	0.976	0.204	0.427	1.123	0.459	1.651	1.080	1.425	2.875
Cut 2	2.596	1.406	1.250	1.382	1.956	1.583	2.715	3.006	2.806	3.385
Cut 3	3.059	2.410	2.535	2.747	3.081	2.860	3.650	4.216	3.724	4.118
Overall Difference in Unadjusted Latent Depression Means between UK and India[b] = .457 p < .001										

Note: ** p < .001, * p < .05; p < .10; [a] Constrained to 1 for model identification; [b] Represents UK mean relative to Indian mean

6.4.2.4 Comparison of common estimation elements in the mean shift and free threshold AV models

The different elements of the two AV models will now be presented in parallel and compared, starting with commonalities between the models and then moving onto their unique elements, including their overall effects on the comparison of latent depression means between countries.

6.4.2.4.1 Item factor loadings

Responses to self-reports and AVs are a function of a latent factor for subject specific bias, while self-reports are additionally a function of the latent depression factor. Each model therefore estimates the loading of each EPDS item onto both the subject specific bias factor and the latent depression factor. Subject-specific bias indicates how much of the variance in ratings for a given item are the product of individual differences in rating style, while the latent depression factor indicates how much of the variance in ratings was due to actual differences in underlying depression.

Model estimates are presented for the mean shift model in Table 6.4.2.4.1a and for the free threshold model in Table 6.4.2.4.1b. In both models, the factor loading (first row in Tables 6.4.2.4.1a/b) of each item to latent depression was significant, indicating that each item contributed significantly to the latent depression factor, and that underlying depression accounted for a significant amount of variation in self-reported depression. Although there were some small differences in magnitude, with estimates from the free threshold model being marginally larger for each item, the pattern of loadings was broadly similar. Items 7 ("I have been so unhappy that I have had difficulty sleeping"), 8 ("I have felt sad or miserable") and 9 ("I have been so unhappy that I have been crying") made the largest contribution to the underlying depression factor. These results are consistent with the observation that these items carry the most overt references to sadness, which is likely to be the emotion most commonly associated with depression. In contrast, items 1 ("I have been able to laugh and see the funny side of things") and 2 ("I have looked forward with enjoyment to things") made the smallest contributions. These are the only two positively framed items in the scale and have a more abstract connection to postnatal depression, so again, this finding was not unexpected. Item 6 ("Things have been getting on top of me") also had a relatively low factor loading and, while negatively framed, also is fairly abstract and may not contribute as strongly to postnatal depression as more concretely connected items.

There were clearer differences between the models with regards to the subject-specific bias factor (second row in Tables 6.4.2.4.1a/b). While in the mean shift model, each item loaded significantly on this factor, only three items (1, 2 and 6) had significant factor loadings in the free threshold model. This difference is to be expected, as the free threshold model allows for a more nuanced adjustment of ratings by allowing the thresholds to vary between countries for each item.

There are a number of reasons why items 1, 2 and 6 may have retained significant subject specific bias factor loadings in the free threshold model. Firstly, as discussed above, items 1 and 2 are reversed and this may cause confusion and introduce individual variation in ratings. Secondly, these three items contain the most complex phrasing, and this may provoke variation in how individuals respond to the items. Finally, these items also appear to have the most complex response items. Taking one option from each item as an example, responses include “Definitely not so much now” (item 1), reads “Rather less than I used to” (item 2), and “Yes, sometimes I haven’t been coping as well as usual” (item 6), compared to response items such as “Hardly ever” (item 4) and “Yes, quite often” (item 7) in other items. These differences in phrasing complexity may have introduced more individual variation in ratings in both samples, but particularly in the BCHADS cohort. Although WHO guidelines were followed in the translation of the EPDS and every effort was made to minimise differences between the English and Kannada versions, limitations inherent to the EPDS (discussed in Section 5.6.4) and the need to maintain the equivalence of the scale items may have resulted in increased item complexity after being translated into Kannada.

Despite differences in the magnitude of the subject-specific bias factor loading in the two models, it should be noted that in both models the latent depression factor loading was larger than the subject specific bias factor loading in almost every item, and generally, differences were substantial. Overall, this indicates that underlying depression accounted for substantially more variation in ratings than did individual variations in rating style. The main exception to this is item 2, which showed a higher subject-specific bias factor loading in the free threshold model and a marginally lower subject-specific bias factor in the mean shift model. This is consistent with the mean AV ratings for item 2 in the Indian cohort which showed no clear trend in line with expectations regarding vignette severity. This suggests that the combination of the issues outlined in the previous paragraph may have been particularly problematic for Indian respondents.

Table 6.4.2.4.1a: Summary statistics from mean shift model, showing the shared factor loadings for each item, the shared individual vignette and item response threshold coefficients, the mean difference between countries in AV ratings and the adjusted mean difference in self-reported depression between the UK and India.

Coefficients	**EPDS Item**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Latent Depression Factor	.613**	.425**	.758**	.909**	.918**	.656**	1.121**	1.415**	1.146**	.847**
Subject-Specific Bias Factor	-.337**	-.321**	.350**	.254*	.316**	.187*	.283*	.190†	.271*	.237*
Vignette A	2.99**	2.90**	3.05**	2.90**	4.02**	2.58**	4.12**	4.46**	4.44**	5.66**
Vignette B	2.81**	2.65**	1.78**	2.76**	4.01**	2.48**	4.23**	2.97**	3.96**	5.16**
Vignette C	2.28**	2.16**	2.24**	2.66**	3.04**	2.02**	3.29**	3.19**	3.43**	5.10**
Vignette D	1.91**	1.92**	1.52**	1.84**	2.14**	1.72**	3.03**	2.24**	2.52**	4.13**
Vignette E	1.31**	1.67**	1.11**	1.04**	1.89**	1.17**	2.30**	1.52**	2.01**	4.04**
Vignette F	0.60**	0.74**	-0.11	-0.60**	0.22	-0.47**	-0.12	-0.86**	0.87**	1.64**
Cut 1	1.453	0.599	0.412	0.332	0.892	0.654	1.206	.604	0.729	3.141
Cut 2	2.269	1.262	1.682	1.373	1.950	1.654	2.304	2.485	2.225	4.986
Cut 3	3.690	2.205	3.244	3.028	3.542	2.853	3.544	3.956	3.798	5.155
Country Mean Difference[a]	.059	-.681**	.955**	.502**	.421**	.728**	.323*	.598**	-.053	.913**
Overall Difference in Adjusted Latent Depression Means[a] = .021 p = .802										

Note: ** p < .001, * p < .05; p < .10; [a] Represents UK mean relative to Indian mean (reference group)

6.4.2.4.2 Vignette coefficients

The position of each AV on the underlying depression scale was fixed between countries and estimated relative to the Indian mean rating (0). Individual estimates are presented in Tables 6.4.2.4.1a and 6.4.2.4.1b for the mean shift and free threshold models respectively (labelled Vignette A-F). The pattern and ordering of vignette estimates was identical between models, although there were some small variations in estimate magnitude. These estimates were also ordered in the same way as the mean ratings, indicating that the models estimated AV ratings as intended. As shown by the profusion of positive and significant vignette coefficients in both Table 6.4.2.4.1a and Table 6.4.2.4.1b, vignette ratings were almost all significantly more severe than the mean self-report rating. Only the least severe vignettes for items 3, 4, 6, 7 and 8 were rated as less severe than the mean self-report. This is not unexpected as the majority of mothers in a community sample will not report high levels of depressive symptoms. The aim of the current approach was to provide vignettes which evenly divided the response scale. While this appears to have been broadly achieved, in hindsight, a better approach may have been to provide more vignettes that pertained to the lower levels of latent depression, as this would have provided a higher resolution of DIF at this more heavily used end of the scale.

Table 6.4.2.4.1b: Summary statistics from free threshold AV model, showing the shared factor loadings for each item, the shared individual vignette coefficients, the unique item response threshold coefficients, the adjusted mean difference in self-reported depression between the UK and India.

Coefficient	EPDS Item									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Shared										
Latent Depression Factor	.456**	.296**	.777**	.938**	.970**	.756**	1.251**	1.593**	1.293**	.892**
Subject-Specific Bias Factor	.376**	.448**	.078†	.031	.111†	.242**	.046	-.032	.007	.051
Vignette A[b]	3.05**	3.05**	2.85**	2.74**	3.91**	2.89**	4.27**	4.64**	4.56**	5.43**
Vignette B[b]	2.85**	2.79**	1.61**	2.61**	3.93**	2.87**	4.34**	2.99**	4.09**	4.91**
Vignette C[b]	2.25**	2.31**	2.04**	2.50**	2.90**	2.21**	3.31**	3.24**	3.46**	4.83**
Vignette D[b]	1.84**	1.96**	1.30**	1.63**	1.97**	1.76**	2.96**	2.15**	2.37**	3.84**
Vignette E[b]	1.24**	1.72**	0.85**	0.80**	1.67**	1.10**	2.16**	1.31**	1.77**	3.70**
Vignette F[b]	0.48**	0.67**	-0.46**	-0.94**	-0.11	-0.95**	-0.59*	-1.37**	0.42*	1.34**
Range	2.57	2.38	3.31**	3.68	4.04	3.84	4.93	6.01	4.14	4.09
India										
Cut 1[b]	1.567	.794	.517	.349	.855	1.066	1.388	.776	.990	2.819
Cut 2[b]	2.060	1.151	1.451	1.299	1.877	1.584	2.342	2.484	2.007	3.775
Cut 3[b]	3.14	2.107	2.520	2.596	3.171	2.260	3.192	3.466	3.478	4.850
Range	1.574	1.314	2.003	2.207	2.316	1.193	1.804	2.69	2.489	2.03
UK										
Cut 1[b]	1.105	1.101	-1.144	-.692	-.002	-.849	.197	-.918	-.008	1.978
Cut 2[b]	2.264	2.229	.401	.405	1.113	.816	1.560	1.376	2.213	2.603
Cut 3[b]	4.056	3.206	2.364	2.517	3.156	3.003	3.572	3.780	4.206	3.954
Range	2.952	2.105	3.508	3.209	3.158	3.852	3.375	4.698	4.214	1.976
Overall Difference in Adjusted Latent Depression Means[a] = -.451 p < .001										

Note: ** p<.001; * p<.05, † p <.1; a - represents UK mean relative to Indian mean; b – estimated relative to Indian mean.

6.4.2.5 Comparison of distinct estimation elements in the mean shift and free threshold models.

After following a common approach for the estimation of factor loadings and vignette coefficients, the two models then diverge in terms of how DIF is estimated and adjusted for. In the mean shift model, thresholds between countries were shared, and a mean difference in AV ratings for each item between countries was estimated, with India as the reference group. The overall mean difference between countries was then applied to the latent depression factor before individual scores are extracted. Accordingly, country mean difference in Table 3 indicates the severity of UK ratings of AVs relative to Indian ratings of AVs. A significant positive coefficient indicates that UK mean ratings were significantly higher than Indian ratings, while a significant negative coefficient indicates that UK mean ratings were significantly lower. In contrast, in the free threshold model the positions of the AVs were fixed between countries but response scale thresholds were allowed to vary in each item between countries, thereby capturing bias directly from the threshold locations. The findings from these two approaches are described below. The mean shift model is discussed first and is examined in relation to the raw mean ratings reported in Table 1. Next, results from the free threshold model are described for each item and compared to results from the mean shift model.

6.4.2.6 Mean shift model adjustment

The overall difference in latent depression means between countries after the mean shift adjustment was non-significant (*d* = .02, *p* .802; Table 6.4.2.4.1a bottom row), indicating that there was no country difference in reported rates of postnatal depression at 12 months after accounting for DIF. This contrasts markedly with the conclusion drawn from comparison of raw self report scores in the naïve model. Yet in the mean shift model, there were clear significant mean differences, indicative of bias, in ratings given to vignettes between countries for eight out of ten EPDS items (Table 3; “Country Mean Difference”) suggesting that adjusting for bias was necessary. In seven of these cases, UK respondents rated vignettes as more severe than Indian respondents. The largest differences were in items 3 and 10. In one case, item 2, UK respondents rated the vignettes as less severe than Indian respondents. There was no evidence of bias in the remaining two items, 1 and 9.

The country mean differences in the mean shift AV model are broadly supported by the descriptive statistics presented in section 6.4.2.1. Mean differences in AV ratings between UK and Indian mothers presented in Table 6.4.2.1 show that each set of vignettes was rated, overall, more severely by UK mothers, other than the set for item 2, broadly

mirroring the pattern observed in the mean shift model. However, there are some inconsistencies as well. For example, item 1 which showed a similar country mean difference to item 9 in the AV model showed a more substantial mean difference in the raw scores. Further, item 8 showed a relatively moderate country mean difference in the AV model but showed the largest mean difference in the raw scores.

Further limitations to the mean shift model became apparent when the raw mean ratings were explored separately for Indian and UK respondents. Figure 6.4.2.1b shows that although mean ratings were typically higher for UK respondents for more severe vignettes, the magnitude of bias actually appeared to decrease as the vignettes became less severe and shifted direction in all but one of the least severe vignettes. The mean shift model may be sensitive to these changes to a small degree, in that the country mean difference will decrease as a result, but it is not sensitive to the degree that it allows for different levels or directions of bias for different vignette severities. If this is the case, then the mean shift model may be underestimating bias at the higher end of the underlying depression scale and overestimating bias at the lower end of the of the underlying depression scale.

6.4.2.7 Free threshold model adjustment

The overall difference in latent depression means between countries after free threshold adjustment was significant and negative (d = -.45, p < .001; Table 6.4.2.4.1b bottom row), indicating a medium sized country difference, with UK respondents reporting lower levels of postnatal depression at 12 months after correcting for DIF. This constitutes a reversal of the overall direction of country difference observed in the naïve model. There was a relatively consistent pattern of differences in the item response thresholds applied to AVs between India and the UK. This model is best examined graphically and so model estimates are presented below for each item in the form of figures. To facilitate comparison between the two approaches to modelling DIF, the estimates for the mean shift model are also presented as figures and discussed where relevant. For the sake of simplicity, the mean shift model will be referred to as model 1, and the free threshold model referred to as model 2. First, a detailed comparison of item 1 is presented for illustrative purposes, followed by more simplified figures and brief descriptions of each of the remaining items.

Of primary interest in the figures is the location of the thresholds for each item and how they differ between countries and models. The response thresholds ($\tau_1 - \tau_3$) are represented by approximate normal distribution curves in order to demonstrate that the likelihood of endorsing a given response changes based on the level of the underlying depression and to

facilitate comparison with similar figures presented by King et al., (2004). However, these curves are for illustration only and do not represent precise estimated likelihoods. In each of the model 1 figures, it can be seen that the position of the thresholds for each country are shared, reflecting that they are fixed at the same point in the mean shift model, and that an overall bias factor estimated from the mean country difference in AV ratings for each item is applied uniformly to UK self-reports in order to adjust for differences in response style. This effectively applies a uniform shift to the UK thresholds in response to the cumulative bias detected from each set of vignettes. In contrast, it can be seen that the positions of the thresholds in the model 2 figures are not shared. This reflects the fact that thresholds were allowed to vary in this model in response to variation in AV ratings. It is this variation in the thresholds that allows UK self-reports to be adjusted so that they are equivalent to Indian self-reports, thus accounting for DIF. As well as providing a clear visual insight into how response styles differ between the two groups, this also provides a clear way to observe differences in the models and to determine which appears to be more sensitive to variation in DIF between countries and items. Additionally, the figures demonstrate how well the AVs break up the responses, or thresholds (cut points), across the distribution and between countries. This gives an indication as to the discriminatory power and overall resolution of each AV set (King et al., 2004; King, 2009). Vignette locations will be discussed in relation to figures for the free threshold model.

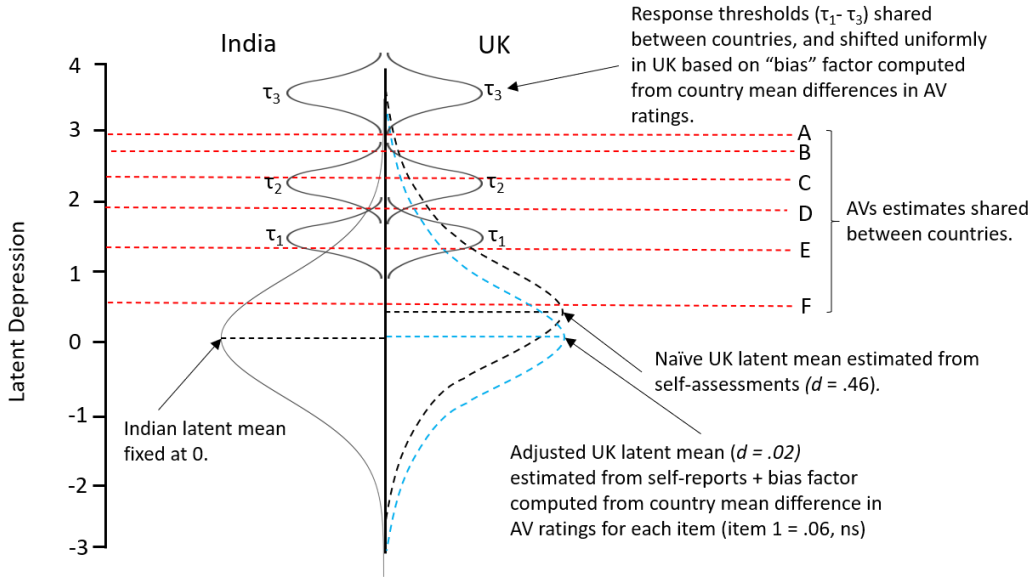
6.4.2.7.1 Item 1: "I have been able to laugh and see the funny side of things"

In model 2 (Figure 6.4.2.7.1b), a lower τ_1 estimate for UK participants indicates that they had less demanding requirements for what constitutes "not quite so much". A similar τ_2 for UK and Indian participants indicates that they had similar requirements for what constitutes "definitely not so much". A higher τ_3 for UK participants indicates that they had more demanding requirements for what constitutes "no not at all". Model 1 (Figure 6.4.2.7.1a) found an overall non-significant mean difference (*d* = .06) in AV ratings between countries. In the larger picture of the mean shift model, this item specific mean difference contributes to the overall mean shift adjustment factor that is then applied uniformly across all items and thresholds. The non-significant mean difference appears to be obscuring the unique and opposite variation in the threshold locations identified for τ_1 and τ_3 in Model 2, meaning that the resultant contribution to the final "mean shift" is neutral, potentially leading to an under-adjustment of scores across all items. This pattern of the mean shift model being insensitive to variations in threshold locations between countries is

repeated in many of the items illustrated below and suggests that, overall, the mean shift model is likely to have under-adjusted self-ratings.

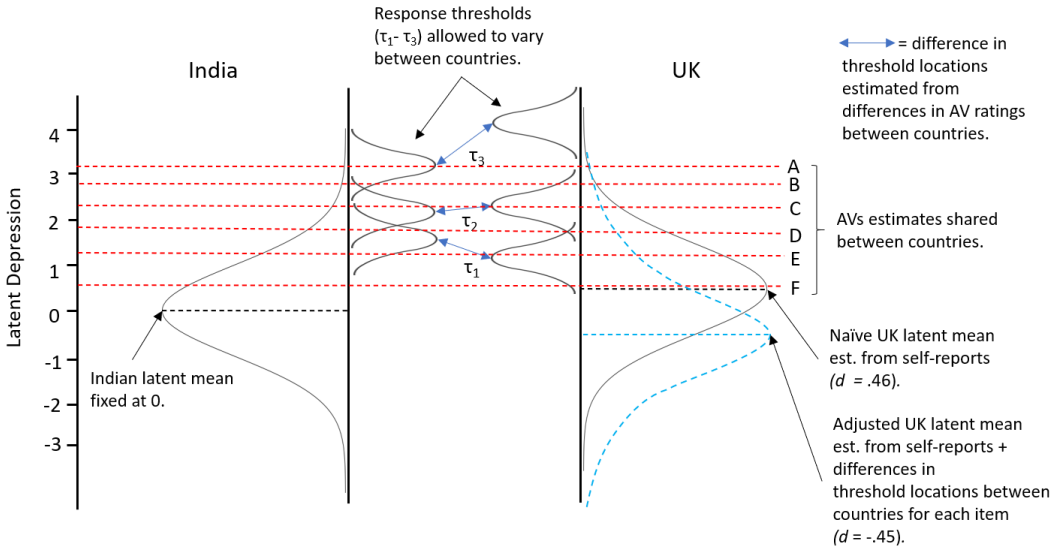
The vignettes in this set (for item 1) were evenly spaced and broke up the thresholds well. This indicates that the vignettes should provide adequate discriminatory power between countries and are reasonably efficient, with no evidence of vignettes clustering together. A possible improvement to this set would be to add a vignette at the higher end of the latent depression scale as there is currently none located above threshold 3. Vignettes were also all located above the mean latent depression rating for the Indian sample. This is repeated in several of the vignettes below. While not ideal, it is understandable given the low levels of depression present in community samples.

Figure 6.4.2.7.1a: Model 1 (mean shift) estimates for EPDS item 1



Note – all parameters estimated relative to Indian latent mean (0).

Figure 6.4.2.7.1b: Model 2 (free threshold) estimates for EPDS item 1

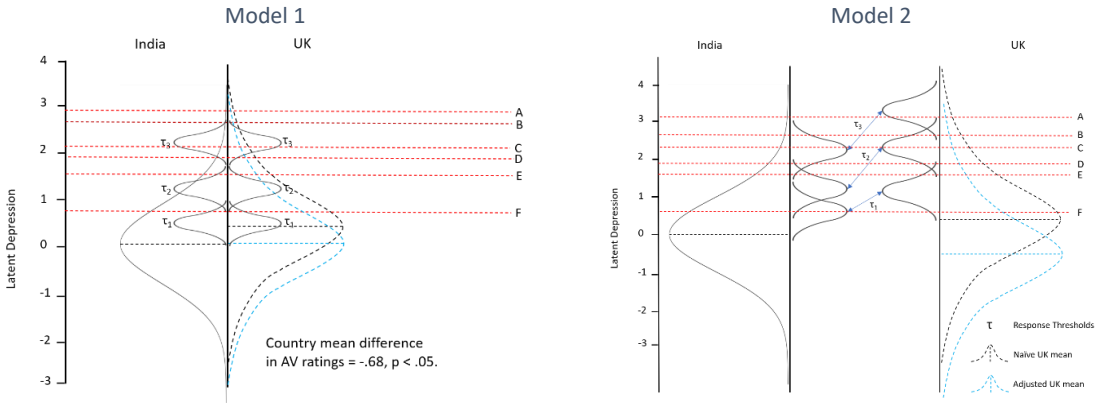


Note – all parameters estimated relative to Indian latent mean (0).

6.4.2.7.2 Item 2: "I have looked forward with enjoyment to things"

Higher threshold estimates in model 2 (Figure 6.4.2.7.2) for τ_1, τ_2 and τ_3 indicate that UK participants had higher requirements for every category in this response set. Unlike most items, this corresponds to the finding in the model 1 that UK participants rated the AVs as significantly more severe than Indian participants.

Figure 6.4.2.7.2 AV model estimates for item 2

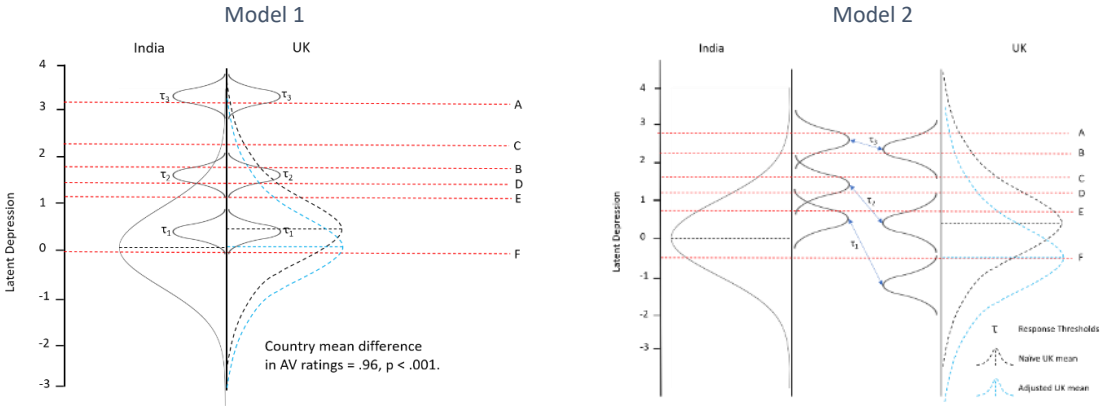


Note - Vignettes A-E are slightly skewed to higher end, but relatively evenly spread and divide threshold well. An additional AV between E & F may increase discriminatory power.

6.4.2.7.3 Item 3: "I have blamed myself unnecessarily when things went wrong"

Lower threshold estimates in model 2 (Figure 6.4.2.7.3) for τ_1, τ_2 and τ_3 indicate that UK participants had lower requirements for every category in this response set. The uniform correction applied by model 1 is not sensitive to potential changes in the magnitude of threshold variation at the higher and lower ends of the scale.

Figure 6.4.2.7.3 AV model estimates for item 3

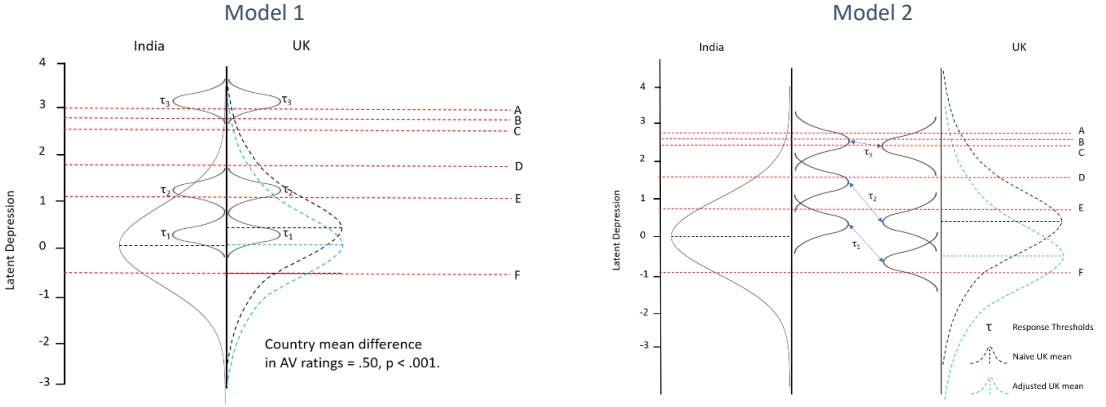


Note - Vignettes A-E are slightly skewed to higher end, but evenly spread and divide threshold well.

6.4.2.7.4 Item 4: "I have been anxious or worried for no good reason"

Lower threshold estimates in model 2 (Figure 6.4.2.7.4) for τ_1, τ_2 and τ_3 indicated that UK participants had lower requirements for each category in this response set.

Figure 6.4.2.7.4 AV model estimates for item 4

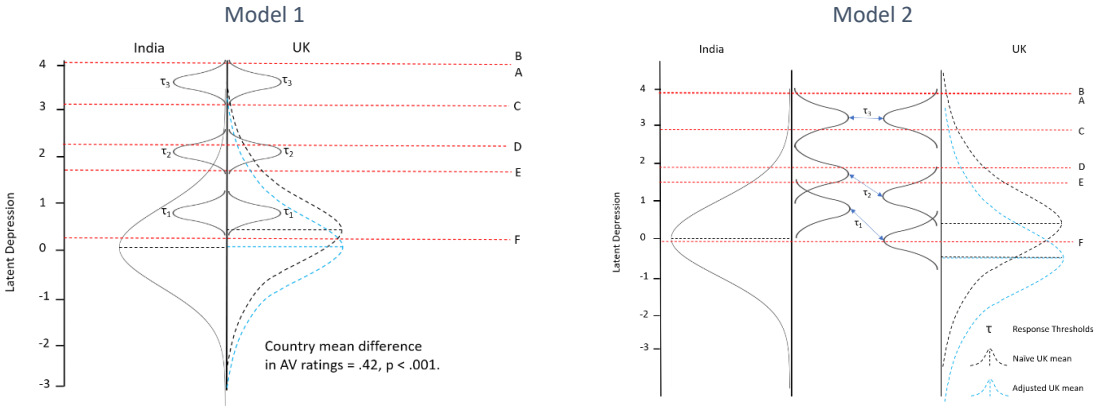


Note - Vignettes A-C are clustered at the top of the scale, and there is a large gap between E & F. Set could be improved by removing AV from the high cluster and adding AV between E & F.

6.4.2.7.5 Item 5: "I have felt scared or panicky for no very good reason"

Lower threshold estimates in model 2 (Figure 6.4.2.7.5) for τ_1 and τ_2 indicate that UK participants had lower requirements for what constitutes for "no, not much" and "yes, sometimes". The similar τ_3 threshold indicates that UK and Indian participants had similar requirements for what constitutes "yes, quite a lot".

Figure 6.4.2.7.5 AV model estimates for item 5

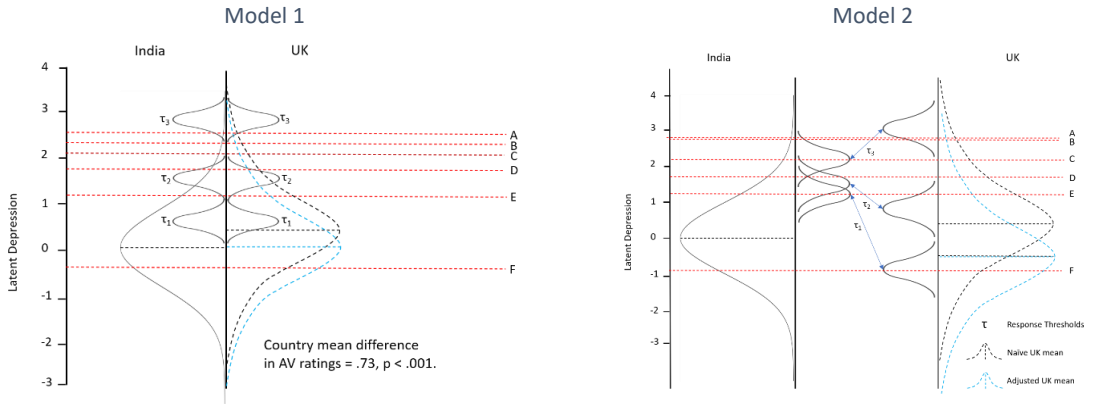


Note - AVs A & B almost identical and so offer little unique information, but remaining vignettes divide distribution of thresholds well. Set could be improved by removing A or B and adding AV between E & F.

6.4.2.7.6 Item 6: "Things have been getting on top of me"

Lower threshold estimates in model 2 (Figure 6.4.2.7.6) for τ_1 and τ_2 indicate that UK participants had lower requirements for what constitutes for "no, most of the time I have coped quite well" and "yes, sometimes I haven't been coping as well as usual". A higher estimate for τ_3 indicates that UK participants had higher requirements for what constitutes "Yes, most of the time I haven't been able to cope at all".

Figure 6.4.2.7.6 AV model estimates for item 6

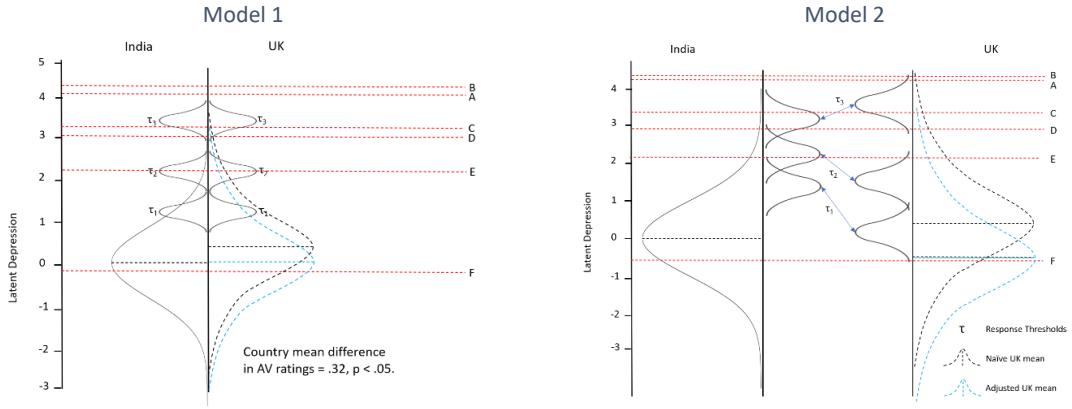


Note - AVs A & B almost identical and so offer little unique information, but remaining vignettes divide distribution of thresholds well. Set could be improved by removing A or B and adding AV between E & F.

6.4.2.7.7 Item 7: "I have been so unhappy that I have had difficulty sleeping"

Lower threshold estimates (Figure 6.4.2.7.7) for τ_1 and τ_2 indicate that UK participants had lower requirements for what constitutes for "not very often" and "yes, sometimes". Higher τ_3 indicates that UK participants had higher requirements for what constitutes "Yes, most of the time".

Figure 6.4.2.7.7 AV model estimates for item 7

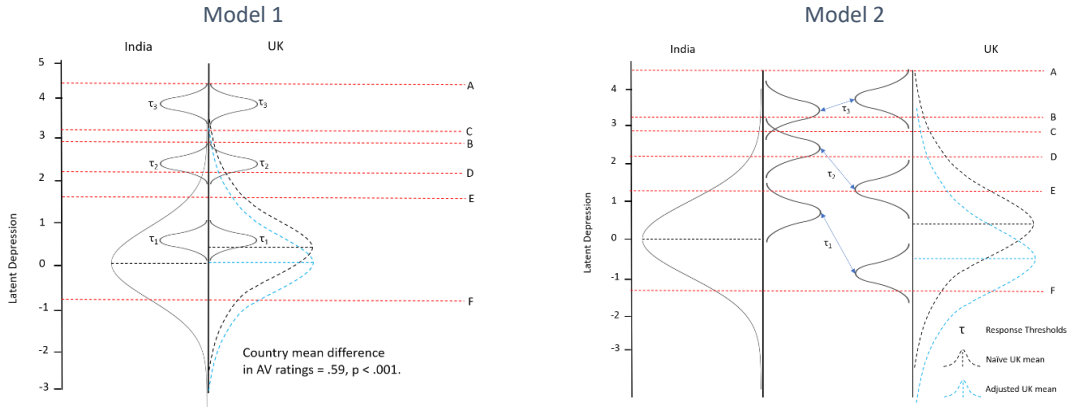


Note - Note: AVs A & B almost identical and so offer little unique information, and remaining vignettes are more skewed toward the top of the scale creating a large gap between E & F. Set could be improved by removing A or B and adding multiple AVs between E & F.

6.4.2.7.8 Item 8: "I have felt sad or miserable"

Lower threshold estimates in model 2 (Figure 6.4.2.7.8) for τ_1 and τ_2 indicate that UK participants had lower requirements for what constitutes for "not very often" and "yes, quite often", respectively. A higher estimate for τ_3 indicates that UK participants had higher requirements for what constitutes "yes, most of the time".

Figure 6.4.2.7.8 AV model estimates for item 8

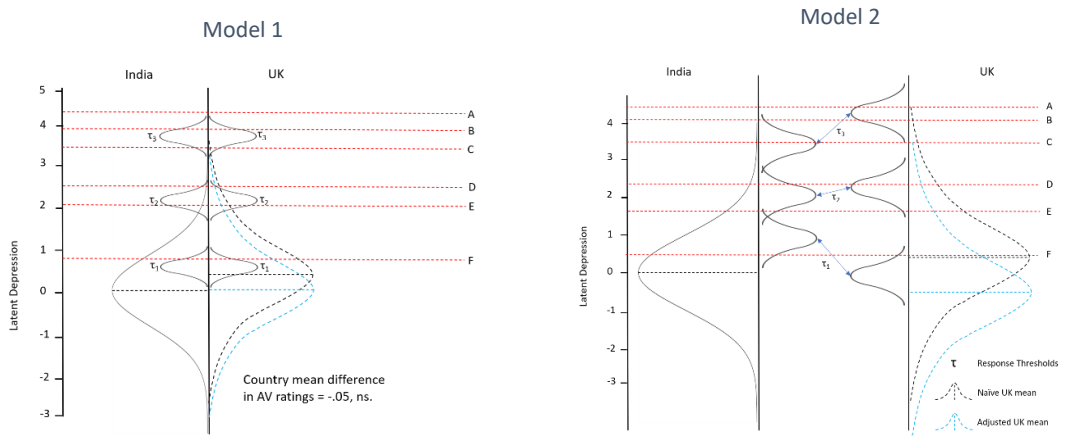


Note - AVs are generally evenly spaced and divide up the thresholds well. The large gap between E & F appears to be the result of lower ratings for F, rather than the upward skew seen in other items.

6.4.2.7.9 Item 9: "I have been so unhappy that I have been crying"

A lower threshold estimate in model 2 (Figure 6.4.2.7.9) for τ_1 indicates that UK participants had lower requirements for what constitutes "only occasionally". However higher estimates for τ_2 and τ_3 indicate that UK participants had higher requirements for what constitutes "Yes, quite often" and "yes, most of the time", respectively. Model 1 (Fig 3.17) found no evidence of a significant overall mean difference in country ratings of AVs, and this appears to be a reflection of variation in different cut-point locations "cancelling" each other out.

Figure 6.4.2.7.9 AV model estimates for item 9

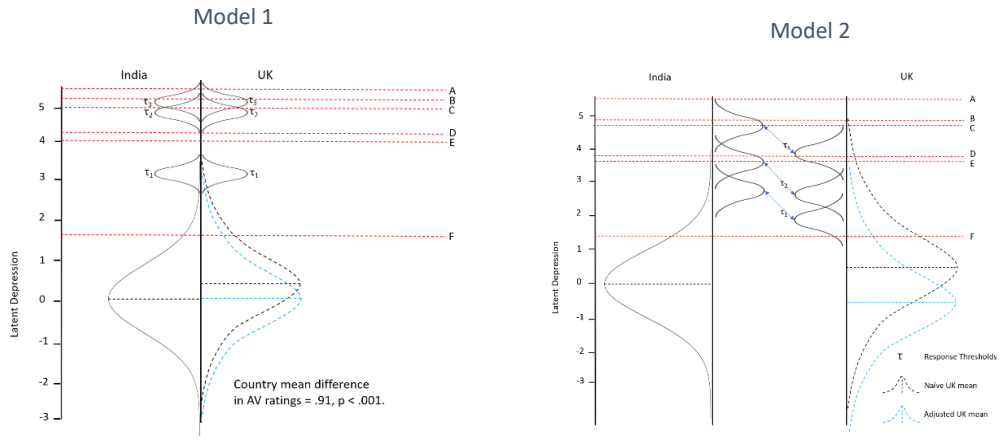


Note - AVs were evenly spaced and broke up the distribution of the thresholds well.

6.4.2.7.10 Item 10: "The thought of harming myself has occurred to me"

Lower threshold estimates in model 2 (Figure 6.4.2.7.10) for τ_1, τ_2 and τ_3 indicate that UK participants had lower requirements for each response category in relation to this item. The mean country difference estimated by model 1 (Fig 3.19) appears to be relatively accurate as it indicates that UK mothers rate AVs significantly more severely.

Figure 6.4.2.7.10 AV model estimates for item 10



Note - AVs A-E are all substantially skewed to the top end of the scale. While thresholds are also clustered at the high end of the scale, indicating the AVs probably did discriminate adequately between them, it may have been helpful to have additional AVs toward the bottom of the scale. Additionally, AVs B & C and D & E share very similar locations and so will be providing little unique information.

6.4.2.8 Summary of free threshold results

For the majority of items model estimates indicated that UK participants used lower thresholds for τ_1 and τ_2, which represent the cut-points between ratings of 0/1 and 1/2, and used similar or higher thresholds for τ_3, which represents the cut-point between ratings of 2/3. This indicates that, generally, UK participants had lower standards for ratings of 1 or 2 but higher standards for ratings of 3. In contrast, Indian participants generally had higher standards for ratings 1 or 2 but lower standards for ratings of 3. This suggests that Indian mothers may have been more likely to rate vignettes as 0 or 3, while UK mothers may have been more likely to rate vignettes as 1 or 2.

While, intuitively, a higher τ_3 threshold may appear to suggest that UK mothers would rate more severe vignettes as less severe than their Indian counterparts, this is not what was observed in the mean ratings (Figure 2). One reason for this may be the dispersion of the lower thresholds. As can be seen from the figures, the Indian τ_2 and τ_1 thresholds were typically higher than the corresponding UK thresholds, meaning that a larger proportion of the underlying latent depression scale fell under these lower thresholds in India than it did in the UK. Consequently, this may have driven down the ratings of vignettes for Indian mothers, resulting in the trends seen in the mean ratings. The differences between countries for τ_1 and τ_2 (where the UK had lower thresholds) are larger than for τ_3 (where the UK had higher thresholds), adding further to the larger proportion of the distribution that fell into lower categories for Indian respondents relative to UK mothers. However, it is unclear how this pattern of thresholds fits with the trend for Indian mothers to rate the least severe vignettes as more severe than UK mothers. It is possible that this is the result

of an inadequate number of vignettes at the lower end of the latent depression scale, thereby not providing enough discriminatory power to accurately model DIF at this point.

There were two clear exceptions to the trend discussed above. The first is item 2, in which all thresholds were higher for UK participants, indicating that UK participants had higher standards for each category in this item. In this case, it would be expected that vignettes would be rated lower by UK participants and this was reflected in the mean ratings, as well as the mean shift model. The other exception is item 10, in which all thresholds were substantially lower for UK participants, indicating that UK participants had lower standards for each of these categories. Again, given the uniform distribution of the thresholds, it would be expected that vignettes would be rated as more severe by UK participants, and again, this is clearly shown in the mean ratings and further supported by the mean shift model.

6.4.2.9 How well do the AVs divide up response thresholds?

On the whole, there was a fairly even spread of vignettes and they appeared to divide up the threshold distribution very well. A small issue is that vignette distribution tended to be slightly skewed toward the higher end of the distribution in a number of items, and markedly so in items 7 and 10. However, in most cases the AVs were able to discriminate between thresholds in the two countries. There is fairly consistent evidence that the more severe vignettes, particularly vignettes A and B, were located at very similar positions on the underlying scale. Little additional information is provided by vignettes that are located closely together and so in many of these cases one of the two vignettes could have been dropped to increase efficiency without sacrificing discriminatory power. There was also a tendency for quite a large gap between vignettes E and F and it is possible that this resulted in insufficient discriminatory power to fully detect DIF at the lower end of the distribution. Thus, the method may have been further improved by adding a vignette with a severity pitched between these two to improve the resolution of the correction score at this point. This may be particularly important as this represents the point on the scale where the majority of respondents rated themselves. Overall, though, the vignettes appeared to be more than adequate to provide the discriminatory power required to detect and adjust for DIF between the two countries.

6.4.2.10 How do the two AV models compare?

Overall, the free threshold model appears to be more sensitive to the way in which DIF functions in the present sample. This approach was able to detect variation in thresholds

within items, with all items showing different levels of variation between categories, and the majority showing different directions in the variation of threshold between categories. For example, many showed lower threshold estimates for τ_1 and τ_2 in the UK sample and threshold estimates for τ_3 that were either similar between countries or higher in the UK, and this is supported by the mean ratings as discussed above. The estimates provided by the mean shift model were not sensitive to these variations and therefore obscured them in presenting a simple shift in one direction or the other. Although this was appropriate for items 2 and 10, it was not appropriate for the other items and applying this uniform correction may have resulted in over-adjusting or under-adjusting the thresholds for certain categories. Therefore, on this evidence, it appears that the free threshold model is a more sensitive approach with the current data.

6.4.2.11 Model comparison using Bayesian information criterion

The two AV models were not nested so it was not possible to directly compare them using an overall likelihood-ratio test. Instead Bayesian information criteria (BIC) were used to compare models. Models differed in the number of parameters that were freely estimated, with model 1 estimating 130 parameters and model 2 estimating 155 parameters. The BIC for model 1 = 41,763.583, while the BIC for model 2 = 40,492.490. This indicates that model 2 provided a better fit to the data, despite estimating a higher number of parameters.

6.4.3 Exploring the accuracy of the AV models.

A series of random effects logistic regression models were then tested for each item to further explore which approach to DIF adjustment is most suitable for the data. The first two models treated the vignettes as a set of dummy variables coded A (most severe) to F (least severe), with the first (Model 1a) investigating the main effect of country on vignette ratings and the second (Model 1b) investigating the interaction between the effect of country and vignette severity on ratings (Table 6.4.3a).

The main effect of country on vignette ratings followed the same pattern of observed effects in the mean shift model, both in terms of magnitude and direction of differences, indicating that this model detected genuine effects. However, although model 1a supported the findings of the mean shift model, estimates from Model 1b suggested it was not providing a full picture of underlying bias. For the individual vignettes, a significant effect indicates that the mean difference between countries for that vignette was different to the mean difference between countries for the reference vignette (Vignette A), while the 5DF test indicates significant differences across the whole set. Table 6.4.3a shows at least

two significant interactions between country and vignette in each item set and that the overall test of interaction was significant for all items. This suggests that bias was not shared equally among vignettes, and the trend for more significant effects in vignettes D-F indicates a possible split in how bias functions with more and less severe vignettes.

In order to explore this further, a continuous variable representing vignette severity was generated from the vignette dummy variables (0 - most severe to 5 - least severe) and used to test a one degree of freedom interaction effect between country and severity on vignette ratings (Model 2). Results are presented in Table 6.4.3b and indicated that there were significant differences in bias as severity decreased for each item. Estimates suggest that in all items, UK participants gave the more severe vignettes higher ratings than Indian respondents, but that these discrepancies declined as the severity of the vignettes declined. Although not possible to confirm with these estimates, it is possible that this decline may, in some cases, pass through the point of “no bias” and start operating in the other direction. This interpretation is supported by the mean ratings which generally show higher UK ratings for the more severe vignettes but that these differences decline as the vignettes

Table 6.4.3a: Output from Model 1a and 1b - showing main effects of country, and interactive effects of country and vignette on vignette ratings

Coefficient		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Model 1a	**B**	-0.53^{+}	-0.35	-2.03**	-0.27	0.23	-0.31	0.33	-1.99**	-0.68*	-0.96*
	C	-1.60**	-1.14**	-1.28**	-0.51**	-1.66**	-1.20**	-1.21**	-1.53**	-1.57**	-0.86*
Vignette	**D**	-2.13**	-1.54**	-2.84**	-1.77**	-3.29**	-1.90**	-1.55**	-2.79**	-3.33**	-2.69**
	E	-3.24**	-1.93**	-3.64**	-3.00**	-3.65**	-2.73**	-2.80**	-3.79**	-4.18**	-2.80**
	F	-4.76**	-3.61**	-5.93**	-5.63**	-7.08**	-5.19**	-7.01**	-7.06**	-5.71**	-7.56**
Mean Country Difference (UK)		0.36^{+}	-0.78**	0.77**	0.63*	1.18**	0.64*	0.97**	0.98**	-0.37^{+}	2.14**
Model 1b	**B x UK**	-1.36*	-1.36*	-0.74	-0.34	0.34	2.00**	-0.99	-1.91**	-0.86	0.93
	C x UK	-0.87	-1.32*	-0.58	0.23	-0.69	-0.78	-2.23**	-2.34**	-2.47**	-0.94
	D x UK	-0.79	-2.23**	-0.83	-0.93*	-0.83	0.09	-1.76**	-2.24**	-2.88**	-1.36^{+}
Vignette x Country	**E x UK**	-1.14*	-2.05**	-1.93**	-1.25*	-1.16*	-0.24	-1.84**	-2.75**	-3.18**	-2.51**
	F x UK	-3.34**	-5.20**	-4.49**	-4.01**	-3.34**	-2.21**	-5.58**	-6.30**	-5.06**	-3.18**
	5DF Test	X^2=25.3 p<.001	X^2=67.4 p<.001	X^2=62.7 p<.001	X^2=63.3 p<.001	X^2=38.6 p<.001	X^2=43.9 p<.001	X^2=43.4 p<.001	X^2=63.2 p<.001	X^2=73.9 p<.001	X^2=40.7 p<.001

$^{+}$p = .05-.10, *p < .05, **p < .001.

Table 6.4.3b Output from Model 2 - showing main effects of country, severity, and interactive effects of severity (continuous) and country on vignette ratings

Coefficient	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Country (UK)	1.23**	1.07*	2.37**	2.22**	2.92**	1.73**	2.80**	2.73**	2.24**	5.43**
Severity	-0.71**	-0.24**	-0.59**	-0.68**	-0.87**	-0.62**	-0.71**	-0.70**	-0.66**	-0.48**
Country X Severity	-0.34*	-0.70**	-0.64**	-0.63**	-0.68**	-0.45**	-0.74**	-0.72**	-0.89**	-1.22**

$^{+}$p = .05-.10, *p < .05, **p < .005.

grow less severe and that, in most cases, the least severe vignette is actually rated as more severe by Indian respondents.

6.4.3.1 Which model do these results support?

Although the results from model 1a support the findings from the mean shift model and indicate that the country mean differences detected by that model are indeed a genuine reflection of overall mean differences in data, the results from model 1b and model 2 indicate that only considering overall mean differences is likely to obscure more nuanced variation in how response thresholds are utilised in the two groups. The interaction between country differences and vignette severity instead suggests that the free threshold model provides a far more accurate model of DIF in the current sample. Findings regarding differences in bias across vignettes are consistent with findings from the free threshold model, which shows a trend for UK mothers to apply lower values to thresholds T_1 and T_2 than Indian mothers, but similar or higher values to threshold T_3. As discussed previously this pattern of thresholds is, somewhat counter-intuitively, consistent with the conclusion from the continuous interaction model that bias functions in one direction for the most severe vignettes (UK – more severe). Although it is not clear from the free threshold model why the least severe vignettes are rated more highly by Indian respondents, in general, the model is more sensitive to changes in the level of bias across the different thresholds than the mean shift model.

6.4.4 Summary of findings

The two approaches to estimating the AV-adjusted latent depression means produced a shift in the same direction but with different magnitudes. When estimated using the mean shift model, the difference in latent depression means between countries became non-significant (d = .02, p < .80). When estimated using the free threshold model, the difference in latent depression means switched direction and there was a significant and medium-sized negative effect of country (d = -.45, p < .001), indicating that, after adjustment for DIF, UK mothers reported lower levels of postnatal depression than Indian mothers at 12 months. Thus, for both models AV ratings indicated that overall, the apparent higher level of depression amongst UK mothers is a function of DIF and does not reflect differences in actual levels of latent postnatal depression. However, the magnitude of the adjustment is dependent on the modelling approach taken. Comparisons of AV model estimates, model fit statistics, raw data and other exploratory analyses suggests that the free threshold model is better suited than the mean shift model to detecting and adjusting for DIF in the current sample.

6.5 Discussion

The current study investigated the longitudinal and cross-cultural properties of the Edinburgh Postnatal Depression Scale (EPDS) in India and the UK using two advanced statistical methodologies, measurement invariance (MI) testing and anchoring vignettes (AVs). Longitudinal MI testing found evidence of partial metric invariance in India between 8 weeks and 12 months and partial scalar invariance in the UK between 8-weeks and 12-months. Cross-cultural MI testing found no evidence of metric invariance between countries at 8 weeks but did find evidence of partial metric invariance at 12-months. Consequently, in order to make direct comparisons of latent means between countries at 12 months, further information was required regarding the item response thresholds used in each sample in order to adjust responses to a common scale. Anchoring vignettes were therefore used to detect differential item functioning (DIF) in the use of response options.

Due to the exploratory nature of this study and the novelty of the AV approach with the EPDS two different AV models were tested. The adjusted latent means were extracted from each model and compared to the non-adjusted mean depression score, which showed that UK mothers reported significantly higher depressive symptoms. Both AV models resulted in a significant adjustment, with the mean shift model indicating that there was no significant difference between latent means after adjustment, and the free threshold model indicating that UK mothers actually reported significantly lower depressive symptoms following adjustment. Both models therefore indicate that UK mothers have higher expectations for mental health and tend to rate their own symptoms more severely than Indian mothers. However, the free threshold appears to be more sensitive to variations in thresholds used for different response categories and may therefore be producing a more accurate adjustment. Findings from each level of the investigation and their implications for cross-cultural research are now discussed.

6.5.1 EPDS measurement Invariance

Measurement invariance testing was conducted as a precursor to the use of the anchoring vignette analysis, which is the main focus of this study. Accordingly, the results from these analyses will only be discussed briefly to provide context for a more detailed discussion of AV results.

6.5.1.1 Longitudinal measurement invariance

Given the importance of assessing the longitudinal impact of maternal depression during the postnatal period and throughout infancy (NICHD ECRN, 1999) and the tendency in the

literature to compare studies which recruit samples at different points (Martin & Redshaw, 2018), longitudinal invariance was tested between EPDS assessments at 8 weeks and 12 months postnatal, separately in India (BCHADS) and the UK (WCHADS).

In BCHADS, there was evidence of configural invariance and partial metric invariance in 9/10 items, but there was no evidence of scalar invariance. This indicates that the underlying construct being measured by the EPDS is the same across time-points and that most items contribute to the latent depression factor in a similar way at 8 weeks and 12 months. The decision to test invariance in this way was partially driven by the finding that there are distinct EPDS cut-offs at 8 weeks and 6-24 months in the current study for indicating depression caseness on diagnostic interview (see Section 3.2.2.1), and the finding of scalar non-invariance across time points is consistent with those results. Taking these findings together, it is apparent that although PND may represent an equivalent concept, mothers are applying different thresholds to their responses and different levels of the latent trait are required for item endorsement between the two time-points in the perinatal period. It is possible that sociocultural factors relating to expectations during the immediate perinatal period may play a role in influencing how a mother reports her own symptoms. This has important implications for the use of EPDS data across time-points in this setting. Firstly, although the lack of scalar invariance means that the same participant response at each time-point was not equivalent, invalidating direct comparisons of latent means across time-points, the fact that partial metric invariance was achieved means that each item in the scale made an equivalent contribution to the underlying factor (depression) at each timepoint. This means that the distribution of the underlying factor is equivalent between time-points. This is similar to the AV models which constrained the factor loadings of each item to be the same between countries in order to ensure that the distribution underlying latent trait (depression) was the same between countries. This means, for example, that the association between PND assessed using the EPDS at 8 weeks and cognitive development at 24 months is comparable to the association between PND assessed using the EPDS at 12 months and cognitive development at 24 months.

Secondly, and more relevant to the AV analysis, is that given the assumption that participants use response scales in the same way between self-assessments and AV assessments (response consistency), the use of different thresholds means that it may not be appropriate to apply an AV correction score generated at one time-point to EPDS data collected at another. In the present study, the AV data was collected at 12 and 24 months and so it cannot be used to correct the 8-week self-assessment data.

Invariance was stronger in the WCHADS sample, with results indicating full configural and metric invariance, and partial scalar invariance for 6/10 items, allowing for valid comparison of latent means across time-points. Given that the EPDS was originally developed and validated using a UK sample, it is not surprising that the psychometric properties of the measure are stronger in this context. However, significant differences in intercepts for item 4 (I have been anxious or worried for no good reason), item 5 (I have felt scared or panicky for no good reason), item 7 (I have been so unhappy I have had difficulty sleeping) and item 9 (I have been so unhappy that I have been crying) indicate that there are changes during this period which are influencing how UK mothers rate their own depression. While significantly more positive and supportive of invariance than those reported by Cunningham, Brown and Page (2015), these findings support their conclusion that caution is warranted when using the scale in postpartum groups across different time-points.

6.5.1.2 Multi-Group measurement invariance

Milfont and Fischer (2010) state that testing for measurement invariance between cultures is especially important due to biases introduced by sociocultural factors that may affect the meaning each group ascribes to scale items. In the present study, measurement invariance between India and the UK was tested at two time-points to explore the need for AVs. The rationale for testing separately at 8 weeks and 12 months was based on the different cut-off points validated previously and the non-invariance observed in the longitudinal testing. At 8 weeks there was evidence of configural equivalence but only 4/10 items were invariant at the metric level. As this did not meet the standard set out by Vandenberg and Lance (2000) for partial invariance, further levels of invariance were not tested. This indicates that although the construct of depression is perceived in broadly the same way, the indicators (items) were not contributing to the latent variable equivalently between cultures. This is consistent with the validation study reported in chapter 3, which established a cut-off of 3 or greater for probable depression in India, considerably lower than the cut-off of 13 or greater established in the UK (Cox et al., 1986).

At 12 months there was evidence of configural and partial metric invariance in 6/10 items, but not evidence of scalar invariance. Items 1 (I have been able to laugh and see the funny side of things), 2 (I have looked forward with enjoyment to things), 6 (Things have been getting on top of me), and 8 (I have felt sad or miserable) all showed significantly different factor loadings at the metric level. The potential reasons why will be discussed in more detail later on but it is possible that items 1, 2 and 6 differed between countries as they all

contain complex stem questions and response sets, both with phrasing that is specific to the English language. Although every effort was made to translate these items accurately, it may be that this complex and colloquial language lost some of its original meaning in the Kannada EPDS.

What these results do imply is that comparisons of within-country differences and relationships with postnatal depression at 12 months can be compared across countries, but more caution is required in interpreting comparisons at 8 weeks. This has important implications for the current study as it suggests that the results obtained in chapter 3 regarding the relationship between chronic depression and infant cognitive development are comparable to those from a UK context, but the results regarding early postnatal depression may not be as straight forward. This is partially addressed by utilising a cut-off point validated in the current population but further harmonisation work may be required to provide more confidence for comparisons. Beyond this, the lack of scalar invariance at 12 months indicates that the origin, or intercept, of the EPDS scale in one country is shifted relative to the other. This is similar to the example given in Chapter 4 (Section 4.3.3) of the 273 degree difference in scale origin between Kelvin and Celsius, meaning that although the two scales are measuring the same construct (configural invariance) and that a 1 point change on the scale represents an equivalent change in the construct (metric invariance), the scale itself has a different intercept between groups and a score of 5 degrees Kelvin (e.g. India) is not equivalent to a score of 5 degree Celsius (e.g. UK). Therefore, in order to make direct comparisons of latent mean scores between the UK and India, more information is required to discern the origin of each scale so that responses can be adjusted to a common scale. This provides a strong rationale for the use of anchoring vignettes at this time-point.

6.5.2 Summary of anchoring vignette results

As already stated, estimates of the differences in latent depression means between India and the UK were significantly affected by AV adjustment. Prior to adjustment, estimates showed that UK mothers reported significantly higher depressive symptoms than Indian mothers, but following adjustment this difference either became non-significant (mean shift model) or became significantly different in the opposite direction (free threshold model).

6.5.2.1 What do the models tell us about differences in response style between India and the UK?

The shift in the difference between latent means following DIF-adjustment in both models is indicative of a tendency for UK mothers to rate the vignettes and themselves more severely than Indian mothers, suggesting that the two models are detecting bias that functions in a consistent direction. However, the magnitude of the correction is substantially different, with the free threshold model applying a much larger correction. Examination of the thresholds predicted by the free threshold model suggests that the mean shift model may be underestimating bias because it assumes that country specific DIF is affecting all thresholds equally. When thresholds are estimated freely, rather than showing a uniform shift in one direction, the differences in threshold locations between countries appears to vary from one cut-point to another. The dominant pattern is for τ_1 and τ_2, representing the cut-points between ratings of 0/1 and 1/2, to be lower for UK participants, and for threshold 3, representing the cut-point between ratings of 2/3, to be at either a similar level between countries or marginally higher for UK participants. Although this means that the likelihood of a UK mother rating any vignette as 3 (most severe) is lower than that for an Indian mother, it also means that a larger proportion of the latent depression scale is located between the lower thresholds for Indian mothers, leading to an overall decrease in their ratings. Specifically, the largest difference between countries tends to be in the location of τ_1, leading to a far higher likelihood of Indian mothers endorsing a rating of 0. In the mean shift model, where thresholds are assumed to move in a uniform way, the differences in τ_1 are likely to have been attenuated by the smaller, and sometimes reversed, differences in τ_3. This is just one example of how the mean shift approach may be obscuring more significant and nuanced cultural differences in the use of response thresholds and leading to an underestimation of DIF.

Exploratory analyses also indicate stronger support for the free threshold approach with the current sample. The patterns observed in the mean ratings tended to show that UK mothers gave higher ratings for the more severe vignettes, and that these differences decreased as the vignettes became less severe until, in most cases, the least severe vignette was actually rated as more severe by Indian mothers. While differences in mean ratings do not translate directly to differences in thresholds, this trend does indicate that DIF is not constant and appears to be a function, not only of country group, but also of symptom severity. This was further shown in a set of random effects logistic regression analyses of the AVs alone, which revealed significant interactions between symptom

severity and country for each set of vignettes, with estimates showing that bias in favour of more severe UK ratings was strongest in the most severe items and that it decreased with vignette severity. These results suggest that the free threshold model provides a more accurate and sensitive approach for detecting and correcting DIF in EPDS responses between UK and Indian participants.

6.5.2.2 Are the adjusted latent means more accurate?

Another metric of model accuracy is whether the adjusted rates of depression appear to be valid and more closely aligned with expectations regarding the differences between LMICs and HICs. As there is no objective indicator of postnatal depression available in the current study, model accuracy is considered in relation to prevalence estimates in the existing literature. Standing in stark comparison to the unadjusted rates of depression, published meta-analyses drawing from studies that have used a combination of clinically assessments and symptoms questionnaires have consistently found higher rates of postnatal depression in LMICs. Woody, Ferrari, Siskind, Whiteford, and Harris (2017) reported that prevalence was significantly higher in LMICs than HICs (18.7% vs 9.5%), while Shorey et al. (2018) are more specific in reporting prevalence rates in Asia (16%) were significantly higher than those in Europe (8%). Additionally, although not directly comparing different settings, Fisher et al. (2012) found that the mean pooled prevalence of common mental disorders in LMICs was 19.8% and Upadhyay et al. (2017) reported a pooled prevalence of 19% for postnatal depression in India alone. These results clearly support the direction of the adjustment in both AV models in the current study but lend further support to the free threshold model in terms of the magnitude of the adjustment.

A further consideration, however, is that each of these meta-analyses combined prevalence estimates from studies using a mixture of clinical interviews and self-report questionnaires. In accordance with the current study, the latter could be expected to be influenced by DIF and so these comparisons themselves may lack validity. Interestingly three of these meta-analyses reported that prevalence estimates from self-report instruments were actually higher than those from clinical diagnostic instruments (Fisher et al., 2012; Shorey et al., 2018; Upadhyay et al., 2017). Specifically, within India there are two studies which have shown relatively high prevalence rates of depression using locally validated versions of the EPDS. Fernandes et al. (2011) reported a prevalence of 14.4% for prenatal depression using a clinical interview and found that a cut-off of ≥13 on a Kannada version of the EPDS adequately detected probable depression at this level. Additionally, Patel, deSouza, and

Rodrgues (2003) reported a prevalence of postnatal depression of 23% using a validated cut-off of ≥13 for a Konkani version of the EPDS. Although symptom scales generally produce higher prevalence rates than diagnostic interviews, these results do not line up with expectations regarding under-reporting from the current study. Instead, they indicate that not all mothers in India or other LMIC settings are under-reporting depressive symptoms. It is possible that differences in the timing of assessment (e.g. prenatal vs postnatal) or demographic factors (e.g. rural vs urban), as well as differences in methodology, may be contributing to apparent differences in reporting behaviour between those studies and the current sample. This is discussed in more detail below. What is clear is that more research is required to explore the effects of DIF in different LMIC settings.

6.5.2.3 What is driving DIF between UK and India?

In the seminal AV study, King et al. (2004) explain that DIF observed between Mexican and Chinese participants regarding the topic of political efficacy was likely to have been rooted in different expectations in each population as to what it meant to have 'a say in government'. The same may be true in the present study with regards to expectations relating to different depressive symptoms. Variations in the expectations of each population in relation to a given symptom (e.g. 'so unhappy that I have been crying') mean that differing levels of that symptom are required for a given response option (e.g. 'yes, most of the time') to be endorsed in each group. Specifically, the present results indicate that UK participants generally have higher expectations regarding postnatal mental health, meaning that less severe symptoms are required for them to endorse higher ratings of depression in themselves. This finding is consistent with findings from other studies which have shown that participants from HICs typically over-rate their own symptoms of mental and physical health relative to participants from LMICs (Knott et al., 2017a; Molina, 2017).

One reason that has been put forward as an explanation for differences in health expectations is different levels of education, either generally or specifically regarding a particular construct. Sen (2002) illustrated this with an example from within India. The state of Kerala, which has the highest levels of literacy in the country, also has both the highest levels of life expectancy and the highest levels of self-reported morbidity. Meanwhile, states such as Bihar, with relatively low life expectancy and poor medical and education facilities, have the lowest rates of self-reported morbidity. Sen argued that this counter-intuitive trend is the result of individuals in Kerala being in a position to more accurately diagnose and perceive their own illnesses and because expectations of what it means to be

healthy have risen as a result of access to better treatment. This position is supported by within-country AV studies which have found that education level is a significant predictor of DIF in a number of different countries including in HICs (Bago d'Uva et al., 2011; Hinz, 2017), LMICs (Hanandita & Tampubolon, 2016; Molina, 2016; Rossouw, Bago d'Uva, & Van Doorslaer, 2018), and specifically within India (Dasgupta, 2018).

Thus, general differences in levels of education may be driving variations in DIF between the two cohorts examined in the present study. However, more specifically, differences in education and awareness regarding mental health may also be contributing to significantly different internal standards. According to Fisher, de Mello, Izutsu, and Tran (2011) the past five decades have seen a substantial increase in research focused on maternal mental health in HICs. This has followed a decrease in maternal mortality and preceded an increased awareness of the psychological aspects of pregnancy, childbirth, and the postnatal period in these settings. In contrast, the shift in focus from mortality to morbidity has only recently started to gain momentum in resource-constrained settings, meaning that maternal perinatal disorders are still under-recognised and general awareness of the importance of mental health during this period is lower (Fisher et al., 2010; Upadhyay et al., 2017). Although there are many examples of innovative initiatives addressing maternal health throughout LMICs (see, for example, Chandra, Desai, Reddy, Thippeswawmy, & Saraf, 2015) there has so far been little success in the scaling up of these services to meet needs on a national scale (Eaton et al., 2011). According to a recent expert meeting on the subject of maternal mental health in resource-constrained settings, this has resulted in a *"widespread lack of awareness about women's mental health in the perinatal period"* in these contexts (Fisher et al., 2011, p.5). In turn, this relative lack of awareness may be driving the differences in health expectations that are detected by the anchoring vignettes in the present study.

A further possibility is that the increased levels of adversity often observed in LMICs may be contributing to under-reporting of symptoms in the Indian sample. Black et al. (2017) observed that living in a resource constrained setting increases the likelihood of being exposed to multiple adversities. According to Fisher, Rahman, Cabral de Mello, Chandra, and Herrman (2010), mothers in resource-constrained settings are less likely to have completed primary schooling and to have had sufficient sexual and reproductive health education to facilitate autonomous choices regarding family planning. They are also more likely to live in crowded circumstances, to be poorly nourished, to be carrying a coincidental burden of infectious diseases, to be constrained by rigid gender-stereotypes, and are more

likely to be exposed to violence, be devalued and have their rights ignored on the basis of their gender. While the authors point out that these factors may serve as important social determinants of poor perinatal mental health, it is also possible that they may lead to an under-reporting of depressive symptoms relative to high-income settings. Thus, although there are also a number of potential protective factors, such as collectivist principles and the increased prevalence of joint-families and shared-caregiving (discussed in detail in 3.5.6), it is possible that in the context described by Fisher et al. (2010) mothers in India have developed a higher tolerance for the cognitive and emotional symptoms of postnatal depression than do their UK counterparts, and therefore more severe symptoms are required to elicit equivalent self-assessment ratings. If this is the case, the use of scales such as the EPDS without AV correction may be masking the extent of the challenge that maternal perinatal mental health poses in LMIC settings. However, further work is required to establish whether adversities are affecting response style in the hypothesised manner, both generally in resource-constrained settings and specifically in India.

6.5.3 Methodological considerations: novelty, strengths, and challenges in the current study.

Due to the novelty of the current approach, in many ways this project can be considered as exploratory and as such, as much can be learned from the process of development and the performance of the vignettes as can be learned from the actual model output. The following sections will summarise the strengths of the current approach and discuss how this may have impacted the performance of the vignettes, as well as reviewing the key challenges encountered along the way. This will provide a strong platform for the future use of AVs in this field.

6.5.3.1 Approach to development and translation

As detailed in chapter 5, the development of a set of anchoring vignettes for the EPDS was a major part of this project. As part of this process, the author was able to review and draw on the experience and recommendations of the AV literature to create a comprehensive set of guidelines for the development, translation and administration of anchoring vignettes. This preparatory work was key in creating confidence that the vignettes used met the two key measurement assumptions, vignette equivalence and response consistency.

6.5.3.2 Vignette equivalence

Vignette equivalence is the assumption that all participants perceive a given vignette as representing the same level of latent health. This assumption is predicated on two key

elements of development, writing and translation. Vignettes in this project were written using clear, unambiguous language and included descriptions of concrete behaviours and symptoms and the frequency of their occurrence (King et al., 2004). Descriptions of mental health were restricted to universally experienced symptoms such as crying, laughing or tiredness, so as to avoid more abstract descriptors which may have carried distinct cultural meanings (Grol-Prokopczyk, 2018) or have been multidimensional in nature (Vonkova, Bendle, & Papjoanu, 2017). Using a multi-item scale like the EPDS which breaks down a concept like depression into distinct symptoms provides an advantage over single item scales in this regard. This approach reduces ambiguity and minimises the need for participants to add their own subjective interpretations to the vignette, thereby increasing confidence that vignettes are being perceived equivalently across individuals and groups.

The next key in establishing vignette equivalence is taking a functional or conceptual approach to maintaining equivalence in translation. Grol-Prokopczyk et al. (2015) noted that many of the most widely used anchoring vignettes failed strict tests of vignette equivalence and highlighted the lack of attention paid to cultural norms and social practices in translation as key to this. In response, translation into Kannada in the current study was carried out according to WHO guidelines (Menon, Cherkil, Aswathy, Unnikrishnan, & Rajani, 2010) by a committee of psychiatrists, psychologists and researchers who were native to India, had extensive knowledge of the culture, and who were provided with explicit instructions to focus on conceptual equivalence. The author was also based in Bangalore during the translation process and was able to provide a UK cultural perspective on the meaning of different words and concepts. Upon completion of translation, vignettes were piloted, and appropriate modifications were made to the vignettes and to the administration procedures. This comprehensive and stringent approach to development, translation and administration increases confidence of fulfilling the vignette equivalence assumption and resolves past oversights in development identified in the literature (Topp, Heesen, Augustin, Andrees, & Blome, 2020).

Although formal tests of vignette equivalence have been proposed (Bago d'Uva et al., 2011) it has been acknowledged that they are extremely demanding (Grol-Prokopczyk et al., 2015). As a result, many subsequent studies either have not tested this assumption at all (Dasgupta, 2018; Hinz, 2017; Poksinska & Cronemyr, 2017; Roussow, Bago d'Uva, & van Doorslaer, 2018; Topp et al., 2020) or have relied on a variation of the informal rank ordering evaluation proposed by King et al. (2004). Kang and Grol-Prokopczyk (2020) examined rank order by calculating the mean rating of each vignette and compared the

pattern of ratings in each country. Similar patterns were observed in the current study and, alongside the steps taken above, provide reassurance that the latent health in the majority of vignettes was perceived as the same in each country. There were some discrepancies in mean ratings between countries though, particularly for item 2. However, the magnitude of these discrepancies were typically quite small and mostly constrained to more severe vignettes. Overall, the mean ranking comparison between countries suggest that vignettes were perceived in generally the same way and provides support for the VE assumption.

6.5.3.3 Improving confidence in response consistency

Response consistency is the assumption that participants apply the same response thresholds to self-assessments and vignette assessments and relies on both the content of the vignette and the approach to administration. King et al. (2004) suggest that aside from the description of the latent construct, all other vignette content should be geared towards convincing the participant that the character is like them in every way. In the current project, all vignette content was reviewed by local experts from both cultures to ensure that there were no references to activities or behaviours that would seem out of place in each context. Explicit instructions telling participants to consider the character to have the same age and background as them, and to imagine themselves in their position were repeated throughout the administration (Au & Lorgelly, 2014). Piloting indicated that this approach was effective and although it was not possible to perform a strict test of response consistency, Knott et al. (2017a) found that following these same procedures resulted in the AVs passing formal tests. This provides confidence that the response consistency assumption is likely to have been met in the current study.

6.5.3.4 Approach to analysis

This is the first study to use the anchoring vignette methodology to detect and correct for differential item functioning (DIF) in participant responses to the EPDS and, more generally, to our knowledge it the first study to do so for any multi-item scale in the area of mental health research. As far as the author is aware, the only other study to utilise a parametric approach to AV correction for a multi-item scale was Vonkova et al. (2017), who used AVs with a 5-point scale assessing dishonest behaviour in school. There are also a number of studies in the personality literature that apply an AV correction to multi-item scales, but each of these has taken a less powerful non-parametric approach to analysis (Marksteiner, Kuger, & Klieme, 2019; Mottus et al., 2012; He et al., 2017). This is also the first study to

generate a correction score in one sample and apply it to self-assessment ratings in another.

One challenge presented by the EPDS is that it has a unique response set for each item, thereby requiring far more vignettes than in most other studies. Furthermore, because the AV study was embedded in two existing cohort studies, there were some restrictions on the available participants and so the current sample was smaller than in many other studies. The combination of all of these factors meant that a more exploratory stance was taken to analysis. This further adds to the novelty of the current approach and allowed for the strengths and limitations of each approach to be explored by examining results in relation to patterns and effects observed in the raw data. In doing so it became apparent that the AV method is sensitive to the modelling approach taken and that thresholds do vary both between and within items that load onto the same latent construct. This emphasises the need to carefully appraise the estimates provided by AV models rather than assuming that they are representing genuine effects. It also validates the decision in the current study to develop a set of vignettes for each unique response set and to model the thresholds in each separately.

Finally, both AV models included a factor for subject-specific bias. This allowed for variation in each individual's use of response thresholds, resulting in a more accurate correction score. This also allowed comparison of the overall variance in the ratings for each EPDS item that was accounted for by subject-specific bias and by the actual underlying depression factor. Indications were that underlying depression accounted for more variance than individual response style in all items. In the free threshold model, the subject-specific bias factor was only significant for items 1, 2 and 6. Interestingly, these were the three items that failed to show metric invariance between India and the UK at 12 months, indicating that they pose consistent challenges in cross-cultural research. There are a number of reasons why this may have been the case. Firstly, items 1 and 2 are reversed and this may cause confusion and introduce individual variation in ratings. Secondly, these three items contain the most complex phrasing, and this may provoke variation in how individuals respond to the items. Finally, these items also appear to have the most complex response items. These differences in phrasing complexity are likely to introduce individual variation in ratings and could be expected to be further exacerbated after being translated into Kannada. Extra attention may need to be given to these items, either in terms of ensuring that participants are understanding the items and response sets sufficiently, or in future translations of the scale.

6.5.3.5 Key challenges in this context

Primarily, challenges arose in this project as the result of the design of the EPDS. Although only consisting of 10 items, the EPDS has 10 distinct response scales, some of which are very complex and may refer to both past and present states (e.g. *"Yes, sometimes I haven't been coping as well as usual")*. This meant that each item required a distinct set of AVs corresponding to the item content. Many individual AVs also had to be relatively long and complex to make sense in the context of the response options. This made the already difficult and time-intensive process of achieving functional equivalence between two very different cultures even more challenging.

The additional length of the vignettes required by the EPDS also had a large knock-on effect on administration, particularly in India. Following translation into Kannada, the length of many vignettes grew substantially, adding considerably to the administration time and cognitive load placed on participants. This is an important point because the BCHADS sample had relatively low literacy rates and so could not follow the standard procedure of self-administering the vignettes. Instead, AVs were pre-recorded to ensure uniform delivery and played back to participants. Given the length of the vignettes, this method may have increased comprehension errors and heightened the likelihood of vignette mis-ordering. To counter this, participants were instructed to listen to each AV at least twice and extra instructions were added to encourage participants to pay attention to all the relevant details. This form of administration may have also added a sense of time-pressure to Indian participants compared to UK participants who could read and respond in their own time. As a result, it is possible that Indian participants may not have taken as much time to process and evaluate the AV information (Molina, 2017).

6.5.4 Limitations

There are several limitations to the current study. Firstly, the sample that completed the AVs was relatively small. Although it was adequate according to King et al. (2004), other studies have been able to draw from far larger pools of participants. The smaller sample size meant that certain restrictions were placed on analyses, namely that models had to be designed to be as parsimonious as possible and that it was not feasible to include a larger set of relevant covariates that may have also contributed to DIF. One iteration of the mean shift model was performed that included the effects of maternal age, but results did not differ significantly and so this model was not presented as part of the final results. Secondly, a number of the mothers in the BCHADS sample did not understand spoken

Kannada to a sufficient level to be included in the AV assessment, further reducing the available sample size. As these mothers may have been native to other Indian states, their response style may also have differed from those who completed the AVs. Thirdly, there was no objective indicator of postnatal depression available in the sample, meaning that the accuracy of the DIF-corrected scores could only be measured against the expected direction of comparisons in the general literature. These may themselves be subject to DIF and poor cross-cultural validity, although several meta-analyses showed that the results of clinical interviews supported the adjusted results in the current study. Fourth, although examination of the mean AV ratings indicated that the vignette equivalence assumption was broadly met, and evidence from the literature suggests that the response consistency assumption will have been met, formal tests of both assumptions were not conducted. Fifth, examination of the distribution of vignettes suggests that while they were generally well spread out across the distribution of the thresholds, they were skewed slightly toward the higher end of the scale. Specifically, for several items, multiple high severity vignettes clustered together, thereby offering little unique information, while there was a relatively large space between the lowest severity vignette and the next level, indicating that information regarding response style may have been missing toward the lower end of the latent depression scale. Sixth, as discussed in Chapter 5, there were certain limitations inherent to the EPDS that may have impacted on the AVs. Specifically, certain items and response sets were more complex than others, introducing a higher potential for variance in responses. Although every effort was made to maintain equivalence during the translation of the EPDS by the BCHADS study team, it is possible that this complexity was exacerbated after being translated into Kannada.

The BCHADS AV sample consisted of a relatively low-income urban population. As a result, these findings may not generalise to other populations in India, either in rural settings or in other states which vary widely in language, cultural practices and sociodemographic factors (Dasgupta, 2018). Additionally, while there has been a clear adjustment of scores in this case, it should be remembered that AVs are not *"a magic bullet to achieve scalar invariance"* (He et al., 2017, p.13). Both He et al. (2017) and Marksteiner et al. (2019) reported that although the psychometric properties of scales were improved following AV adjustment, they were not improved to the point of achieving full scalar invariance. Given the complexity of the AV adjustment required for the 10 different response sets of the EPDS, the relatively small sample, and the computational time that would be required, this

final step of re-testing measurement invariance following adjustment was not attempted in this study.

6.5.5 Strengths

The key strengths of the current study can be summarised as follows. This is the first study to use the AV method with a multi-item scale in the field of maternal mental health, specifically the EPDS and has shown the approach to be viable, if challenging, in this context. Second, a key advantage of using a multi-item scale like the EPDS is that it breaks down a multi-dimensional concept into concrete and universally experienced symptoms. This should improve the likelihood of meeting the vignette equivalence assumption relative to AV studies which rely on single, multi-dimensional items. A relevant example of this is the single item often used to assess depression, "How much of a problem with feeling sad or depressed" (Mojtabai, 2016; Molina, 2017). By asking a participant to rate the presence of an overarching, multi-dimensional construct such as depression in an AV is ambiguous and could be interpreted differently depending on an individual or societal understanding of what depression entails. Third, an extremely robust and methodical approach was applied in the development, translation and administration of the vignettes, with great care given to applying the knowledge and guidance available in the existing literature. This provides further confidence that the key AV assumptions were met. Fourth, the AVs themselves appeared to generally divide up the distribution of the participant responses well, with a few exceptions. This division is key to determining the discriminatory power of the AVs and therefore provides confidence in their utility. Fifth, two different approaches to modelling DIF detected by AVs were explored and compared in detail, allowing for the approach best suited to the current data to be identified. Sixth, the AV method was supplemented by measurement invariance testing. This provided additional insight as to the functioning of the EPDS between time-points and countries and the results provided a strong rationale for the use of the AV method. Importantly, invariance testing also showed that DIF detected from AV ratings at 12-24 months would not have produced a valid correction score for self-reports at 8 weeks. The lack of scalar invariance between 8 weeks and 12 months indicated that different thresholds are used at these time-points and applying the 12-24 month AVs would have therefore violated the response consistency assumption.

6.5.6 Implications and future directions for research

The current findings stand in broad agreement with previous within-country and cross-cultural AV studies that have found evidence of significant DIF which is dramatically impacting comparisons between different groups in health-related research (Bago D'Uva, Lindeboom, O'Donnell, and Van Doorslaer, 2011; Dasgupta, 2018; Grol-Prokopczyk, Freese, & Hauser, 2011; Hanandita & Tampubolon, 2016; Hinz, 2017; Kang & Grol-Prokopczyk, 2020; Kapteyn, Smith, & Van Soest, 2007; Knott et al., 2017a; Mojtabai, 2016; Molina, 2016, 2017; Rossouw, Bago d'Uva, & Van Doorslaer, 2018) and other fields (King et al., 2004; Mottus et al., 2012; Vonkova et al., 2017; Vonkova, Zumarro, & Hitt, 2018).

The consistency of this evidence adds considerable weight to the argument for the need for data calibration between cultural groups and here this is extended to a multi-item scale of self-reported postnatal depression. Without AV adjustment any direct comparison between the two samples would have led to misleading conclusions. While the EPDS may be a valid predictor of important outcomes within India, as shown by the MI testing at 12 months, the AV findings imply that comparison of raw scores on the EPDS should be used with caution when identifying disparities in rates of postnatal depression between India and the UK. Given that the EPDS was developed and validated in the UK and is known to function well in this setting, and considering the consistent evidence of lower awareness of maternal mental health in LMICs, it can be justifiably argued that the bias seen in the current study should be interpreted as evidence of under-reporting in India, rather than over-reporting in the UK.

In turn this bias may be expected to extend to other groups or countries where awareness of the importance of perinatal mental health is low. Since the EPDS is one the most widely used screeners of perinatal mental health globally and is often a gateway to further diagnosis and service provision, significant under-reporting in LMIC settings is likely to result in many women not receiving adequate care at a time in which they and their children are particularly vulnerable (Cox, 2019; Gibson, McKenzie-McHarg, Shakespeare, Price, & Gray, 2009). However, such assertions must also be considered with the caveat that many studies within India do not demonstrate the same apparent under-reporting of depressive symptoms (Upadhyay et al., 2017). This may be the result of sociodemographic disparities between samples that are driving differences in expectations regarding postnatal mental health, but equally it may be a product of method bias. Specifically, differences in approaches to the translation and the administration of the scale can have a significant

effect on responses given by participants and may be artificially inflating prevalence rates (He & Van de Vijver, 2012). Additionally, it is possible that studies which have found unexpectedly low rates of postnatal depression in these settings are incorrectly assuming that something has gone wrong with their data collection, leading to publication bias, when they are actually observing the product of DIF.

Further research is required, then, to determine if this phenomenon is constrained to the current sample. In this regard, it would be helpful to devise a set of studies with identical approaches to measure development and administration so that this extraneous noise can be removed from the equation, allowing for more accurate comparisons of DIF between settings within India. If it is seen more broadly, it is possible lower mental health awareness and a tendency to under-report symptoms may be an indirect effect of the historically held belief that women in resource constrained settings experience fewer mental health problems because of the lower levels of structured postnatal care provided to them. Although this argument has been refuted as a more systematic approach to mental health research in resource-constrained settings has been adopted, meta-analytic findings of higher prevalence rates in LMIC settings is a relatively recent progression. Thus, the prior erroneous belief regarding the absence of perinatal mental health disorders may have served to slow the progress of improvements to maternal mental health care, precipitating a lack of awareness among women themselves (Fisher et al., 2011). In this case, even where prevalence rates have been found to be higher in LMIC settings than HIC settings, the disparity between contexts may be even larger than previously thought.

Finally, it will be important to return to the analysis conducted in Chapter 3 of this thesis regarding the association between maternal postnatal depression and infant cognitive development, to determine whether the use of the AV corrected score has any bearing on those results. Although the AV method is really designed for cross-cultural comparison, use of the AV corrected scores may still add value to the within-country analysis. Importantly, the AV results showed that response style may vary as a product of the severity of the symptom being rated or experienced. This phenomenon was also observed in a recent study examining the reporting of mental health symptoms by members of the police force in the UK. While there was a general pattern for under-reporting symptoms to an employer relative to a neutral party, the authors observed that those with the most severe symptoms were most likely to under-report symptoms (Marshall et al., 2021). Taken together with the findings from the current study, this suggests that a simple uniform shift of the cut-off value may not be entirely sensitive to the nuances of cross-cultural measurement issues. The AV-

corrected score may therefore provide a more accurate representation of within-sample differences in postnatal depression scores, leading to a more sensitive overall analysis. However, in order to fully replicate the analysis performed in Chapter 3, which compared the impact of exposure to early postnatal and chronic depression, further AVs assessments would be also required to generate a correction score for EPDS scores at 8 weeks post-natal.

6.5.7 Conclusion

In conclusion, the implications of the current findings are potentially twofold. Firstly, differences in response style which indicate that Indian participants are under-reporting depressive symptoms relative to their UK counterparts confirms that direct comparisons between these two populations should not be taken at face value. Any conclusions drawn from such comparisons would be misleading and it is therefore clear that researchers should carefully consider the role of culture in interpreting such data, both from the past and in the future. Consequently, this implies that policy-makers in India and other LMICs should not rely on unadjusted rates of self-reported postnatal depression in their decision making as they may be grossly underestimating the prevalence of mental health disorders. It is important, therefore, to conduct similar studies in other populations within India and other LMICs in order to ascertain whether there is a consistent pattern of under-reporting or whether it is more limited to specific populations. If more widespread, then this will add to the already loud call for better provision of perinatal maternal mental healthcare in these settings and perhaps lend more urgency to the need to raise awareness of this issue both in the professional healthcare service and in the general population. The World Health Organisation (WHO) has devoted considerable energy and resource to scaling up responses to mental health disorders in LMIC settings through the Mental Health Gap Action Programme (mhGAP). However, if the under-reporting of symptoms observed in this study is more widespread, the already worrying statistic that the 76-85% of affected individuals had received no treatment may, in fact, be an underestimation of the scale of the treatment gap (WHO, 2008). Assuming that any systematic under-reporting of symptoms would not be restricted to just postnatal depression, more work is urgently needed to understand the true burden of the whole range of perinatal mental health disorders in LMIC settings and the true scale of the work required to meet this need.

Chapter 7: Executive Summary

This thesis had two distinct but complementary aims. The first, described in Chapters 1-3, was to extend what is known regarding the prospective relationship between maternal depression during infancy and infant cognitive development in India. The second, described in Chapters 4-6, was to evaluate and improve the assessment of postnatal depression in India by using advanced measurement techniques to facilitate a better understanding of how the EPDS functions in this context, relative to how it functions in the UK where it was originally developed and validated. The following sections provide a summary of the overall thesis and the key findings described in the main text.

7.1 Chapters 1-3: Key findings regarding the relationship between maternal postnatal depression and infant cognitive development in Bangalore, India.

7.1.1 Chapter 1 Summary

Chapter 1 presented a comprehensive review of what is known from HIC research regarding the association between postnatal depression and infant cognitive development. This review delineated existing research into two broad categories, with one set of studies focusing the effects of early postnatal depression (i.e., within the first year postnatal), and the other set of studies focusing on the effects of chronic exposure to maternal depression during infancy. Findings regarding the effect of early postnatal depression are relatively inconsistent but there was some indication exposure very early in the postnatal period (<3 months) may be more detrimental than later exposure. Overall, however, there was far more consistent evidence that exposure to chronic depression during infancy posed a significant risk to cognitive development. The review also highlighted evidence that the detrimental effects of maternal depression are transmitted via compromised caregiving, and that this effect was particularly strong in male infants. These findings provided the framework for the current study which focused on comparing the magnitude of the prospective association between early postnatal or chronic depression and infant cognitive development, and the roles of maternal caregiving and infant sex.

7.1.2 Chapter 2 Summary

Chapter 2 then shifted the focus to exploring what is known regarding this relationship in LMIC settings and provided the first systematic review to synthesise findings regarding the associations between maternal postnatal depression and cognitive development during infancy in this context. Nine studies were identified that met the inclusion criteria, the

majority of which only investigated the main effect of maternal depression and not the potentially moderating or mediating effects of other key variables. The review concluded that findings regarding the impact of postnatal depression on cognitive development were inconsistent, with only 4 out of 9 studies finding a significant association, and that more high-quality research was needed in this area that tested main effects and examined moderating or mediating pathways to child cognitive development. These findings further added to the rationale for the current study that had been developed in Chapter 1.

7.1.3 Chapter 3 Summary

After drawing together the framework from Chapters 1 and 2, the study described in Chapter 3 investigated the prospective association between early postnatal and chronic depression throughout infancy and infant cognitive development at 2 years in India. This investigation examined the main effects of depression as well as exploring the role of maternal caregiving and examining whether infant sex contributed to patterns of association. The key finding from this study was that, while there was a significant small effect of early postnatal depression on language subscale scores, following adjustment for covariates neither postnatal nor chronic maternal depression significantly predicted scores on either the cognitive or language subscales of the BSID-III. There was also no evidence of a sex-specific effect of maternal depression, either alone or in interaction with maternal sensitivity.

These findings did not follow the expectation that the multiple, co-occurring adversities generally experienced in LMIC settings, relative to HIC settings, would actually exacerbate the effects of postnatal depression. A number of explanations as to why this was the case were discussed. Firstly, it was noted that the majority of studies which found a significant effect of maternal depression, whether a single early postnatal occurrence or a chronic exposure, defined the exposed group using a clinical diagnosis of depression. While the current study did utilise cut-offs established against a clinical measure, it is known that screening instruments such as the EPDS can lead to the inclusion of women with more mild-moderate symptoms in the exposed group. Thus, in the present study, the depressive symptomology in the early postnatal depressed group and in the mothers scoring more highly on the chronic depression factor may not have been sufficiently severe to impact cognitive development directly or to significantly impair maternal sensitivity.

Two other possibilities were also considered, linked to the socio-cultural context of the BCHADS cohort. Firstly, while elevated levels of risk had been expected to intensify the

effects of maternal depression, it is possible that this heightened adversity may have directly impacted development and caregiving to the extent that the impact of postnatal depression was negligible. This hypothesis is supported by theoretical frameworks such as the Family Stress Model. This framework posits that socioeconomic disadvantage increases familial stress and impairs parental caregiving. The finding that within the sample, SES was a stronger predictor than depression of cognitive development in boys further highlights the role of socioeconomic risk. A second alternative is that the collectivist principles of Indian society and the increased prevalence of joint-families mean that family members are generally more involved in caregiving and able to offer more direct support to mothers than in many HIC settings. This may buffer against the impact of maternal depression on both caregiving and child development. Both of these explanations warrant further investigation.

It is also noteworthy, however, that there was a possible sex-specific effect of maternal sensitivity, with boys of more sensitive mothers showing significantly higher language scores. This extends the pattern of findings in HIC settings that boys are more reliant on external regulation for cognitive development due to a lower capacity for self-regulation. Other sex-specific associations were also noted that sat outside of the main focus of the thesis, namely that boys showed poorer cognitive development when mothers reported the presence of an alternate caregiver, and that multiparity was associated with poorer scores on the cognitive and language subscales for girls. Possible reasons for this are discussed in Chapter 3, but the findings generally point to a possible important interaction between gender and family-level variables in the promotion of infant cognitive development, though this was not directly tested in this study. Alongside the potential role of socio-cultural factors in moderating the association between postnatal depression and infant cognitive development, these findings clearly demonstrate that more detailed investigation regarding the interplay of context-specific factors, such as adversity, family structure and gender attitudes, is needed.

7.2 Chapters 4-6: Key findings regarding the cross-cultural functioning of the EPDS in India and the UK and the feasibility of the anchoring vignette methodology in this setting.

7.2.1 Chapter 4 Summary

The second section of this thesis focused on measurement, specifically with regards to differences in response style between the UK and India. Chapter 4 highlighted the need to examine how cultural differences may lead to measurement bias at different levels. Particular attention was drawn to how group differences in the use of response thresholds

can lead to skewed results. The anchoring vignette methodology was then introduced and a detailed overview of how this technique is able to address this bias, known as differential item functioning (DIF), was provided.

7.2.2 Chapter 5 Summary

Chapter 5 described the development and translation of a new set of anchoring vignettes for use with the EPDS in the present research context. The vignettes were written by the researcher in English and then translated into Kannada. Each step of the development and translation process involved careful review by a team of experts, including clinical psychologists, psychiatrists, and biostatisticians, and by members of the target populations. This was done to ensure that the vignettes met the measurement assumptions implicit in the methodology, namely vignette equivalence and response consistency. This chapter also described the pilot of the anchoring vignettes that was carried in a subgroup of the BCHADS cohort in order to examine the feasibility of the method in that context and to ensure that the vignettes were appropriately worded and translated. The pilot indicated that participants demonstrated an ability to understand and complete the vignette task, that they were rating the vignettes using the same scale and standards that they used to rate themselves, and that vignette content was relevant to participants' lives. A number of limitations to the standard methodology were noted and rectified through minor procedural changes. There were also several limitations relating to the nature of the EPDS itself that were outside of the control of the research team, but that did not critically undermine the approach. Overall, the AV method in general and the vignettes developed as part of this project specifically, were judged to be functioning well and deemed acceptable for use in the present context.

7.2.3 Chapter 6 Summary

Chapter 6 then described an empirical study which utilised the anchoring vignettes (AVs) developed in Chapter 5 to detect and adjust for cross-cultural differences in response style between cohorts in the UK and India. This was the first use of the AV method with the EPDS and, as far as the author is aware, the first time the method has been used to adjust a multi-item scale score in mental health research. In view of this, although the results generated by this study are important in themselves, there is also value in viewing this study as a feasibility test for the AV method in this context. A further novel aspect of the current approach was to generate a bias adjustment factor in one sample which was then utilised to adjust self-reports in another, similar, sample from the same population. In view

of this, an exploratory approach to analysis was taken and bias was estimated using two distinct models: a “mean shift” model, and a “free threshold” model. The estimates from each model were then evaluated to determine which produced a more accurate estimate of DIF.

Prior to adjustment, a mean comparison of the “naïve” self-report data indicated that rates of depressive symptomology were significantly higher in UK mothers than in Indian mothers. However, following AV adjustment, this difference was either attenuated so that there was no significant difference in latent mean scores between countries (Mean Shift Model) or showed a complete reversal so that latent mean scores were significantly higher in the Indian mothers (Free Threshold Model). Thus, in both models, differences were detected between groups in the use of the EPDS response thresholds in the same direction, but the magnitude of the difference was substantially greater when estimated using the free threshold model. Further examination and exploration of the two models indicated that the free threshold model was providing a more accurate estimation of DIF. Specifically, the free threshold model was able to detect variations in bias from one cut-point to another and it was this sensitivity that allowed for a greater correction factor to be estimated. In the absence of an objective diagnostic indicator of postnatal depression, well-established prevalence estimates indicating higher rates postnatal depression in LMIC settings provided further support for the adjustment made by the free threshold model.

The typical pattern of bias revealed that mothers in the UK were more likely to give higher (more severe) ratings to the same symptoms than their Indian counterparts, suggesting that UK participants had higher expectations regarding mental health. This may be the result of higher levels of education regarding mental leading to a higher awareness of mental health in the UK sample. This explanation is consistent with findings from other AV studies which have generally found a higher level of education, either in general or in regard to a specific construct, to lead to lowered response thresholds and higher severity ratings. Alternatively, in the context of increased levels of adversity faced by many women in LMIC settings, it is possible that the Indian participants had a higher tolerance for the cognitive and emotional symptoms of postnatal depression than did the UK participants, and therefore more severe symptoms were required for Indian participants to elicit equivalent self-assessment ratings.

The findings demonstrate the caution required in interpreting self-report data and highlight the value of data harmonisation methods such as anchoring vignettes in cross-cultural

research. This value, however, needs to be weighed against the burden placed on research staff and participants by this approach. Given the time taken to develop and administer the vignettes, and the fact that the correction score generated at the 12-month assessment could not be applied to the 8-week EPDS scores due to apparent longitudinal differences in response style within the BCHADS sample, it could be argued that a diagnostic interview may represent a better investment of resources. Diagnostic interviews are often considered as the gold-standard objective indicator in the assessment of mental health and so their use may reduce the influence of cultural bias, as well as assessing postnatal depression at a level more likely to impact on infant development. However, although they would not require any development work, the use of a clinical research interview does require extensive training and does not offer the same insight into how cultural bias is operating. Additionally, it is not guaranteed that these interviews themselves will not suffer from similar issues relating to bias. In large cohort studies which contain relatively short time-windows for each assessment, a large number of assessors will need to be trained, increasing overall costs and, importantly, increasing the likelihood of poor inter-rater reliability. In contrast, once the AVs have been developed, very little specialist training is required and administration is able to be easily standardised across assessors. In addition, although the initial development of the vignettes is very time consuming, once developed they can potentially be re-used in different cohorts. While further translation work may be required for use in certain samples and the vignettes themselves may need to be administered at multiple time-points, the burden on researchers in terms of development would be greatly lessened. An important next step then in terms of determining the overall utility of this method, is test the feasibility of extending the use of the vignettes developed in this study into other populations. Additionally, as discussed in Chapter 5, the nature and structure of the EPDS (e.g., complex wording, unique response sets) added substantially to the challenges in the development process and these issues may be avoided by utilising a different source tool. Therefore, it is also important that future research explores the feasibility of the AV method with different measurement scales. Finally, given the novelty of this methodology in this context, a relatively cautious approach was necessary which involved using large numbers of vignettes to ensure broad coverage each scale item. A more selective or targeted approach to AV development would both improve the sensitivity of the correction score and reduce the overall burden for researchers and participants.

7.3 Conclusion

Overall, this body of work provides valuable insight into the role of postnatal depression and how it is measured in India. It indicates that apparent low levels of postnatal depression in the BCHADS sample were possibly the product of lower expectations as to what should be considered good mental health. These findings still need to be replicated in other Indian populations but may reflect a lower awareness of mental health amongst women that is a product of the lack of perinatal mental health services available to large parts of the population. This highlights the importance of the work being done by initiatives such as the WHO Mental Health Gap Action Programme to scale up mental health provision in LMIC settings in order to meet the need in these settings. However, while present, postnatal depression may not have the same impact on infant cognitive development as seen in HICs during this period of infancy, emphasising the need to investigate the potentially protective or adverse roles of other biological, nutritional or socio-cultural factors specific to this context and examine their comparative effects during early and later periods of development.

References

Ainsworth, M. D. S., Bell, S. M., & Stayton, D. J. (1974). Infant-mother attachment and social development. In M. P. Richards (Ed.), The introduction of the child into a social world (pp. 99–135). London: Cambridge University Press.

Ahun, M. N., & Côté, S. M. (2018). Maternal depressive symptoms and early childhood cognitive development: a review of putative environmental mediators. *Archives of Women's Mental Health*, *22*(1), 15–24. https://doi.org/10.1007/s00737-018-0870-x

Albers, C. A., & Grieve, A. J. (2016). Test Review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development– Third Edition. San Antonio, TX: Harcourt Assessment. *Journal of Psychoeducational Assessment, 25*(2), 180-190. doi:10.1177/0734282906297199

Ali, N. S., Mahmud, S., Khan, A., & Ali, B. S. (2013). Impact of postpartum anxiety and depression on child's mental development from two peri-urban communities of Karachi, Pakistan: a quasi-experimental study. *BMC Psychiatry, 13*, 274. doi:10.1186/1471-244X-13-274

Au, N., & Lorgelly, P. K. (2014). Anchoring vignettes for health comparisons: An analysis of response consistency. *Quality of Life Research*, *23*(6), 1721–1731. https://doi.org/10.1007/s11136-013-0615-2

Austin, M. P. (2014). Marcé International Society position statement on psychosocial assessment and depression screening in perinatal women. *Best Practice and Research: Clinical Obstetrics and Gynaecology*, *28*(1), 179–187. https://doi.org/10.1016/j.bpobgyn.2013.08.016

Azak, S. (2012). Maternal depression and sex differences shape the infants' trajectories of cognitive development. *Infant Behavior and Development*, *35*(4), 803–814. https://doi.org/10.1016/j.infbeh.2012.07.017

Bagadia, A., & Chandra, P. S. (2017). Starting the conversation - Integrating mental health into maternal health care in India. *The Indian journal of medical research, 145*(3), 267–269. https://doi.org/10.4103/ijmr.IJMR_910_16

Bago d'Uva, T., & Donnell, O. O. (2011). Education-related inequity in healthcare with. *Journal of the Royal Statistical Society*, *174*(3), 639–664.

Bago d'Uva, T. B., Lindeboom, M., O'Donnell, O., & van Doorslaer, E. (2011). Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. *Journal of Human Resources*, *46*(4), 875–906. https://doi.org/10.3368/jhr.46.4.875

Bandoli, G., Coles, C. D., Kable, J. A., Wertelecki, W., Granovska, I. V., Pashtepa, A. O., . . . Cifasd. (2016). Assessing the Independent and Joint Effects of Unmedicated Prenatal Depressive Symptoms and Alcohol Consumption in Pregnancy and Infant Neurodevelopmental Outcomes. *Alcohol Clin Exp Res, 40*(6), 1304-1311. doi:10.1111/acer.13081

Barcellos, S. H., Lleras-muney, A., Carvalho, L. S., & Lleras-muney, A. (2014). Child Gender and Parental Investments In India : Are Boys And. *American Economic Journal: Applied Economics*, *6*(1), 157–189. https://doi.org/10.1257/app.6.1.157.Child

Barclay, K. J. (2015). A within-family analysis of birth order and intelligence using population conscription data on Swedish men. *Intelligence*, *49*, 134–143. https://doi.org/10.1016/j.intell.2014.12.007

Barker D. J. (2004). The developmental origins of chronic adult disease. *Acta paediatrica. Supplement, 93*(446), 26–33. https://doi.org/10.1111/j.1651-2227.2004.tb00236.x

Barnes, J., Gardiner, J., Sutcliffe, A., & Melhuish, E. (2014). The parenting of preschool children by older mothers in the United Kingdom. *European Journal of Developmental Psychology, 11*(4), 397–419. https://doi.org/10.1080/17405629.2013.863728

Baron, E. C., Hanlon, C., Mall, S., Honikman, S., Breuer, E., Kathree, T., … Tomlinson, M. (2016). Maternal mental health in primary care in five low- and middle-income countries: A situational analysis. *BMC Health Services Research*, *16*(1). https://doi.org/10.1186/s12913-016-1291-z

Bateson, P., Gluckman, P., & Hanson, M. (2014). The biology of developmental plasticity and the Predictive Adaptive Response hypothesis. *The Journal of physiology, 592*(11), 2357–2368. https://doi.org/10.1113/jphysiol.2014.271460

Bayley, N. (1993). Bayley scales of infant and development-second edition. San Antonio, TX: The Psychological Corporation.

Bayley, N. (2006). Bayley scales of infant and toddler development-third edition: Administration manual. San Antonio, TX: Harcourt Assessment.

Beck, C.T. (1995). The effects of postpartum depression on maternal-infant interaction: A meta-analysis. *Nursing Research, 44*, 298-304.

Beck, C. T. (1998). *The Effects of Postpartum Depression on Child Development : A Meta-analysis*. *Archives of Psychiatric Nursing, 12*(1), 12–20.

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Arch Gen Psychiatry, 4*, 561-571. doi:10.1001/archpsyc.1961.01710120031004

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: differential susceptibility to environmental influences. *Psychological bulletin, 135*(6), 885–908. https://doi.org/10.1037/a0017376

Bernard, K., Nissim, G., Vaccaro, S., Harris, J. L., & Lindhiem, O. (2018). Association between maternal depression and maternal sensitivity from birth to 12 months: A meta-analysis. *Attachment and Human Development*, *20*(6), 578–599. https://doi.org/10.1080/14616734.2018.1430839

Bhang, S. Y., Ha, E., Park, H., Ha, M., Hong, Y. C., Kim, B. N., . . . Kim, Y. (2016). Maternal Stress and Depressive Symptoms and Infant Development at Six Months: the Mothers and Children's Environmental Health (MOCEH) Prospective Study. *J Korean Med Sci, 31*(6), 843-851. doi:10.3346/jkms.2016.31.6.843

Bjelland, I., Lie, S. A., Dahl, A. A., Mykletun, A., Stordal, E., & Kraemer, H. C. (2009). A dimensional versus a categorical approach to diagnosis: anxiety and depression in the HUNT 2 study. *International journal of methods in psychiatric research, 18*(2), 128–137. https://doi.org/10.1002/mpr.284

Bjerkedal, T., Kristensen, P., Skjeret, G. A., & Brevik, J. I. (2007). Intelligence test scores and birth order among young Norwegian men (conscripts) analyzed within and between families. *Intelligence*, *35*(5), 503–514. https://doi.org/10.1016/j.intell.2007.01.004

Black, M. M., Baqui, A. H., Zaman, K., Mcnary, S. W., Le, K., El Arifeen, S., … Black, R. E. (2007). Depressive symptoms among rural Bangladeshi mothers: Implications for infant development. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(8), 764–772. https://doi.org/10.1111/j.1469-7610.2007.01752.x

Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., … Grantham-McGregor, S. (2017). Early childhood development coming of age: science through the life course. *The Lancet*, *389*(10064), 77–90. https://doi.org/10.1016/S0140-6736(16)31389-7

Blaga, O. M., Shaddy, D. J., Anderson, C. J., Kannass, K. N., Little, T. D., & Colombo, J. (2009). Structure and continuity of intellectual development in early childhood. *Intelligence*, *37*(1), 106–113. https://doi.org/10.1016/j.intell.2008.09.003

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. Developmental Neuropsychology, 28, 561–571. http://dx.doi.org/10.1207/s15326942dn2802_1

Bornstein, M. H., Mash, C., Arterberry, M. E., & Manian, N. (2012). Object perception in 5-month-old infants of clinically depressed and nondepressed mothers. *Infant Behavior and Development*, *35*(1), 150–157. https://doi.org/10.1016/j.infbeh.2011.07.008

Bouchard, T. J., Jr, & McGue, M. (1981). Familial studies of intelligence: a review. *Science, 212*(4498), 1055–1059. https://doi.org/10.1126/science.7195071

Boyd, R. C., Le, H. N., & Somberg, R. (2005). Review of screening instruments for postpartum depression. *Archives of Women's Mental Health*, *8*(3), 141–153. https://doi.org/10.1007/s00737-005-0096-6

Breen, M. S., Wingo, A. P., Koen, N., Donald, K. A., Nicol, M., Zar, H. J., . . . Stein, D. J. (2018). Gene expression in cord blood links genetic risk for neurodevelopmental disorders with maternal psychological distress and adverse childhood outcomes. *Brain Behav Immun, 73*(May), 320-330. doi:10.1016/j.bbi.2018.05.016

Brennan, P. A., Hammen, C., Andersen, M. J., Bor, W., Najman, J. M., & Williams, G. M. (2000). Chronicity, severity, and timing of maternal depressive symptoms: relationships with child outcomes at age 5. *Developmental Psychology*, *36*(6), 759–766. https://doi.org/10.1037/0012-1649.36.6.759

Breton, E. (2019), Modernization and Household Composition in India, 1983–2009. *Population and Development Review, 45*, 739-766. doi:10.1111/padr.12293

Bridgett, D. J., Gartstein, M. A., Putnam, S. P., McKay, T., Iddins, E., Robertson, C., … Rittmueller, A. (2009). Maternal and contextual influences and the effect of temperament development during infancy on parenting in toddlerhood. *Infant Behavior and Development*, *32*(1), 103–116. https://doi.org/10.1016/j.infbeh.2008.10.007

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., … Bhutta, Z. A. (2017). Nurturing care: promoting early childhood development. *The Lancet*, *389*(10064), 91–102. https://doi.org/10.1016/S0140-6736(16)31390-3

Brooks-Gunn, J., Han, W. J., & Waldfogel, J. (2002). Maternal employment and child

cognitive outcomes in the first three years of life: The NICHD study of early child care. *Child Development*, *73*(4), 1052–1072. https://doi.org/10.1111/1467-8624.00457

Buckley, Jack. 2008. Survey Context Effects in Anchoring Vignettes. Working paper retrieved from http://polmeth.wustl.edu/workingpapers.php.

Burger, M., Hoosain, M., Einspieler, C., Unger, M., & Niehaus, D. (2020). Maternal perinatal mental health and infant and toddler neurodevelopment - Evidence from low and middle-income countries. A systematic review. Journal of Affective Disorders, 268, 158-172. doi:https://doi.org/10.1016/j.jad.2020.03.023

Byrne, B. M., Muthèn, B., & Shavelson, R. J. (1989). Testing the equivalence of factor covariance and mean structure: The Issue of Partial Measurement Invariance. *Psychological Bulletin, 105*(3), 456–466.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*(2), 155–175. https://doi.org/10.1177/0022022102250225

Campbell, S. B., Matestic, P., von Stauffenberg, C., Mohan, R., & Kirchner, T. (2007). Trajectories of Maternal Depressive Symptoms, Maternal Sensitivity, and Children's Functioning at School Entry. *Developmental Psychology*, *43*(5), 1202–1215. https://doi.org/10.1037/0012-1649.43.5.1202

Canino, G., Lewis-Fernandez, R., & Bravo, M. (1997). Methodological Challenges in Cross-Cultural Mental Health Research. *Transcultural Psychiatry, 34*(2), 163–184. https://doi.org/10.1177/136346159703400201

Cecil, C. A. M., Pickles, A., Hill, J., & Sharp, H. (2017, October 2). The Joint Contributions of Maternal Depression and Infant Negative Emotionality to Maternal Sensitivity to Infant Distress. https://doi.org/10.31234/osf.io/7pshj

Chadda, R. K., & Deb, K. S. (2013). Indian family systems, collectivistic society and psychotherapy. *Indian Journal of Psychiatry*, *55*(SPEC. SUPPL.), 299–309. https://doi.org/10.4103/0019-5545.105555

Chandra, P. S., Desai, G., Reddy, D., Thippeswamy, H., & Saraf, G. (2015). The establishment of a mother-baby inpatient psychiatry unit in India: Adaptation of a Western model to meet local cultural and resource needs. *Indian Journal of Psychiatry*, *57*(3), 290–294. https://doi.org/10.4103/0019-5545.166621

Cheah, C.S.L. (2016), Commentary: Charting Future Directions for Research on Asian American Child Development. *Child Dev, 87,* 1055-1060. doi:10.1111/cdev.12580

Chen, F. F. (2008). What Happens If We Compare Chopsticks With Forks? The Impact of Making Inappropriate Comparisons in Cross-Cultural Research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Chen, Y., & Baram, T. Z. (2016). Toward understanding how early-life stress reprograms cognitive and emotional brain networks. *Neuropsychopharmacology*, *41*(1), 197–206. https://doi.org/10.1038/npp.2015.181

Cheung, H. S., & Elliott, J. M. (2016). Measuring Maternal Sensitivity: Cultural Variations in the Measurement of Emotional Availability. *Child development, 87*(3), 898–915. https://doi.org/10.1111/cdev.12519.

Cognitive Development. (2020). *Guide for Authors.* Retrieved from

https://www.elsevier.com/journals/cognitive-development/0885-2014/guide-for-authors.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research Using Structural Equations Modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187–212. https://doi.org/10.1177/0022022100031002003

Conger, R. D., & Donnellan, M. B. (2007). An interactionist perspective on the socioeconomic context of human development. *Annual Review of Psychology*, *58*, 175–199. https://doi.org/10.1146/annurev.psych.58.110405.085551

Conroy, S., Pariante, C. M., Marks, M. N., Davies, H. A., Farrelly, S., Schacht, R., & Moran, P. (2012). Maternal psychopathology and infant development at 18 months: The impact of maternal personality disorder and depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(1), 51–61. https://doi.org/10.1016/j.jaac.2011.10.007

Cornish, A. M., McMahon, C. A., Ungerer, J. A., Barnett, B., Kowalenko, N., & Tennant, C. (2005). Postnatal depression and infant cognitive and motor development in the second postnatal year: The impact of depression chronicity and infant gender. *Infant Behavior and Development*, *28*(4), 407–417. https://doi.org/10.1016/j.infbeh.2005.03.004

Corrado, L., & Weeks, M. (2010). Identification strategies in survey response using vignettes. Cambridge Working Papers in Economics. Faculty of Economics, University of Cambridge. Retrieved from https://ideas.repec.org/p/cam/camdae/1031.html.

Cox, J. (2019). Thirty years with the Edinburgh Postnatal Depression Scale: Voices from the past and recommendations for the future. *British Journal of Psychiatry*, *214*(3), 127–129. https://doi.org/10.1192/bjp.2018.245

Cox, M., & Crnic, K. (2003). Qualitative ratings for parent–child interaction at 3–15 months of age. Unpublished Coding Manual.

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of Postnatal Depression: Development of the 10-item Edinburgh Postnatal Depression scale. *British Journal of Psychiatry*, *150*, 782–786. https://doi.org/10.1192/bjp.150.6.782

Crane, M., Rissel, C., Greaves, S., & Gebel, K. (2016). Correcting bias in self-rated quality of life: an application of anchoring vignettes and ordinal regression models to better understand QoL differences across commuting modes. *Quality of Life Research*, *25*(2), 257–266. https://doi.org/10.1007/s11136-015-1090-8

Crockenberg, S. C. (2003). Rescuing the Baby from the Bathwater: How Gender and Temperament (May) Influence How Child Care Affects Child Development. *Child Development*, *74*(4), 1034–1038. https://doi.org/10.1111/1467-8624.00585

Crockenberg, S. C., & Leerkes, E. M. (2003a). Parental acceptance, postpartum depression, and maternal sensitivity: Mediating and moderating processes. *Journal of Family Psychology*, *17*(1), 80–93. https://doi.org/10.1037/0893-3200.17.1.80

Crockenberg, S., & Leerkes, E. (2003b). Infant negative emotionality, caregiving, and family relationships. In *Children's Influence on Family Dynamics: The Neglected Side of Family Relationships* (pp. 57–78). https://doi.org/10.4324/9781410607430

Cunningham, N. K., Brown, P. M., & Page, A. C. (2015). Does the Edinburgh Postnatal Depression Scale measure the same constructs across time?. *Archives of women's*

mental health, 18(6), 793–804. https://doi.org/10.1007/s00737-014-0485-9

Daelmans, B., Darmstadt, G. L., Lombardi, J., Black, M. M., Britto, P. R., Lye, S., … Richter, L. M. (2017). Early childhood development: the foundation of sustainable development. *The Lancet*, *389*(10064), 9–11. https://doi.org/10.1016/S0140-6736(16)31659-2

Dasgupta, A. (2018). Systematic measurement error in self-reported health: is anchoring vignettes the way out? *IZA Journal of Development and Migration*, *8*(1). https://doi.org/10.1186/s40176-018-0120-z

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., … Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*(10), 996–1005. https://doi.org/10.1038/mp.2011.85

Deák, G.O. (2014). Interrelations of language and cognitive development. Encyclopedia of Language Development (pp. 284-291). P. Brooks & V. Kampe, Eds. SAGE

Deave, T., Heron, J., Evans, J., & Emond, A. (2008). The impact of maternal depression in pregnancy on early child development. *BJOG, 115*(8), 1043-1051. doi:10.1111/j.1471-0528.2008.01752.x

Derogatis, L. R., & Savitz, K. L. (1999). The SCL-90-R, Brief Symptom Inventory, and Matching Clinical Rating Scales. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 679-724). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Di Florio, A., Putnam, K., Altemus, M., Apter, G., Bergink, V., Bilszta, J., . . . Meltzer-Brody, S. (2017). The impact of education, country, race and ethnicity on the self-report of postpartum depression using the Edinburgh Postnatal Depression Scale. *Psychol Med, 47*(5), 787-799. doi:10.1017/s0033291716002087

Donald, K. A., Wedderburn, C. J., Barnett, W., Nhapi, R. T., Rehman, A. M., Stadler, J. A. M., … Stein, D. J. (2019). Risk and protective factors for child development: An observational South African birth cohort. *PLoS Medicine*, *16*(9), 1–20. https://doi.org/10.1371/journal.pmed.1002920

Doyle, O., Harmon, C. P., Heckman, J. J., & Tremblay, R. E. (2009). Investing in early human development: Timing and economic efficiency. *Economics and Human Biology*, *7*(1), 1–6. https://doi.org/10.1016/j.ehb.2009.01.002

Dua, T., Tomlinson, M., Tablante, E., Britto, P., Yousfzai, A., Daelmans, B., & Darmstadt, G. L. (2016). Global research priorities to accelerate early child development in the sustainable development era. *The Lancet Global Health*, *4*(12), e887–e889. https://doi.org/10.1016/S2214-109X(16)30218-2

Dunn, L. M. (1965). Peabody picture vocabulary test. Circle Pines, MN: American Guidance Service.

Dutt, R., Raker, C., & Anderson, B. L. (2015). Ethnic variations in cervical cytokine concentrations and vaginal flora during pregnancy. *Am J Reprod Immunol, 73*(2), 141-150. doi:10.1111/aji.12291

Eaton, J., McCay, L., Semrau, M., Chatterjee, S., Baingana, F., Araya, R., Ntulo, C., Thornicroft, G., & Saxena, S. (2011). Scale up of services for mental health in low-income and middle-income countries. *Lancet, 378*(9802), 1592–1603. https://doi.org/10.1016/S0140-6736(11)60891-X

Eberhard-Gran, M., Eskild, A., Tambs, K., Opjordsmoen, S., & Samuelsen, S. O. (2001). Review of validation studies of the Edinburgh postnatal depression scale. *Acta Psychiatrica Scandinavica*, *104*(4), 243–249. https://doi.org/10.1034/j.1600-0447.2001.00187.x

Ekmekci, H., Yavuz-Muren, H. M., Emmen, R. A. G., Mesman, J., van IJzendoorn, M. H., Yagmurlu, B., & Malda, M. (2015). Professionals' and Mothers' Beliefs About Maternal Sensitivity Across Cultures: Toward Effective Interventions in Multicultural Societies. *Journal of Child and Family Studies*, *24*(5), 1295–1306. https://doi.org/10.1007/s10826-014-9937-0

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, *132*(1), 33–72. https://doi.org/10.1037/0033-2909.132.1.33

Erkut S. (2010). Developing Multiple Language Versions of Instruments for Intercultural Research. *Child development perspectives, 4*(1), 19–24. https://doi.org/10.1111/j.1750-8606.2009.00111.x

Evans, J., Melotti, R., Heron, J., Ramchandani, P., Wiles, N., Murray, L., & Stein, A. (2012). The timing of maternal depressive symptoms and child cognitive development: A longitudinal study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*(6), 632–640. https://doi.org/10.1111/j.1469-7610.2011.02513.x

Fagan, J. F., Holland, C. R., & Wheeler, K. (2007). The prediction, from infancy, of adult IQ and achievement. *Intelligence*, *35*(3), 225–231. https://doi.org/10.1016/j.intell.2006.07.007

Familiar, I., Collins, S. M., Sikorskii, A., Ruisenor-Escudero, H., Natamba, B., Bangirana, P., . . . Young, S. L. (2018). Quality of Caregiving is Positively Associated With Neurodevelopment During the First Year of Life Among HIV-Exposed Uninfected Children in Uganda. *J Acquir Immune Defic Syndr, 77*(3), 235-242. doi:10.1097/QAI.0000000000001599

Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 cohort. *Economica*, *70*(277), 73–97. https://doi.org/10.1111/1468-0335.t01-1-00272

Feldman, R., & Eidelman, A. I. (2009). Biological and environmental initial conditions shape the trajectories of cognitive and social-emotional development across the first years of life. *Dev Sci, 12*(1), 194-200. doi:10.1111/j.1467-7687.2008.00761.x

Fergusson, D. M., & Lynskey, M. T. (1993). Maternal age and cognitive and behavioural outcomes in middle childhood. *Paediatric and Perinatal Epidemiology*, *7*(1), 77–91. https://doi.org/10.1111/j.1365-3016.1993.tb00604.x

Fergusson, David M., & Woodward, L. J. (1999). Maternal age and educational and psychosocial outcomes in early adulthood. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(3), 479–489. https://doi.org/10.1017/S0021963098003722

Fernandes, M. C., Srinivasan, K., Stein, A. L., Menezes, G., Sumithra, R., & Ramchandani, P. G. (2011). Assessing prenatal depression in the rural developing world: a comparison of two screening measures. *Archives of women's mental health, 14*(3), 209–216. https://doi.org/10.1007/s00737-010-0190-2

Field, T. (2010). Postpartum depression effects on early interactions, parenting, and safety

practices: A review. *Infant Behavior and Development*, *33*(1), 1–6. https://doi.org/10.1016/j.infbeh.2009.10.005

First, M. B., & Gibbon, M. (2004). The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II). In M. J. Hilsenroth & D. L. Segal (Eds.), Comprehensive handbook of psychological assessment, Vol. 2. Personality assessment (pp. 134-143). Hoboken, NJ, US: John Wiley & Sons Inc.

Fisher, J. R. W., Cabral de Mello, M. C., Izutsu, T., & Tran, T. (2011). The Ha Noi Expert Statement: recognition of maternal mental health in resource-constrained settings is essential for achieving the Millennium Development Goals. *International Journal of Mental Health Systems, 5*(1), 2. doi:10.1186/1752-4458-5-2

Fisher, J., Cabral de Mello, M., Patel, V., Rahman, A., Tran, T., Holton, S., & Holmes, W. (2012). Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review. Bulletin of the World Health Organization, 90(2), 139G–149G. https://doi.org/10.2471/BLT.11.091850

Fisher, J., Tran, T., La, B. T., Kriitmaa, K., Rosenthal, D., & Tran, T. (2010). Common perinatal mental disorders in northern Viet Nam: community prevalence and health care use. *Bulletin of the World Health Organization, 88*(10), 737–745. https://doi.org/10.2471/BLT.09.067066

Fisher, J., Rahman, A., Cabral de Mello, M., Chandra, P., & Herrman, H. (2010). Mental health of parents and infant health and development in resourceconstrained settings: evidence gaps and implications for facilitating 'good-enough parenting' in the twenty-first century world. In *Parenthood and mental health: A bridge between infant and adult psychiatry.* Edited by: Tyano S, Keren M, Herrman H, Cox J. Oxford: John Wiley and Sons; 2010:429-42.

Fox, N. A., Almas, A. N., Degnan, K. A., Nelson, C. A., & Zeanah, C. H. (2011). The effects of severe psychosocial deprivation and foster care intervention on cognitive development at 8 years of age: Findings from the Bucharest Early Intervention Project. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(9), 919–928. https://doi.org/10.1111/j.1469-7610.2010.02355.x

Galler, J. R., Harrison, R. H., Ramsey, F., Forde, V., & Butler, S. C. (2000). Maternal depressive symptoms affect infant cognitive development in Barbados. *J Child Psychol Psychiatry, 41*(6), 747-757. doi:10.1017/S0021963099005910

Ganjekar, S., Thekkethayyil, A. V., & Chandra, P. S. (2020). Perinatal mental health around the world: priorities for research and service development in India. *BJPsych International*, *17*(1), 2–5. https://doi.org/10.1192/bji.2019.26

Garman, E. C., Cois, A., Tomlinson, M., Rotheram-Borus, M. J., & Lund, C. (2019). Course of perinatal depressive symptoms among South African women: associations with child outcomes at 18 and 36 months. *Soc Psychiatry Psychiatr Epidemiol, 54*(9), 1111-1123. doi:10.1007/s00127-019-01665-2

Gelaye, B., Rondon, M. B., Araya, R., & Williams, M. A. (2016). Maternal depression 1 Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. *The Lancet Psychiatry*, *3*(10), 973–982. https://doi.org/10.1016/S2215-0366(16)30284-X

Gengler, J. J., & Mitchell, J. S. (2018). A Hard Test of Individual Heterogeneity in Response Scale Usage: Evidence from Qatar. *International Journal of Public Opinion Research*, *30*(1), 102–124. https://doi.org/10.1093/ijpor/edw025

Gibson, J., McKenzie-Mcharg, K., Shakespeare, J., Price, J., & Gray, R. (2009). A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. *Acta Psychiatrica Scandinavica*, *119*(5), 350–364. https://doi.org/10.1111/j.1600-0447.2009.01363.x

Girault, J. B., Langworthy, B. W., Goldman, B. D., Stephens, R. L., Cornea, E., Steven Reznick, J., … Gilmore, J. H. (2018). The predictive value of developmental assessments at 1 and 2 for intelligence quotients at 6. *Intelligence*, *68*, 58–65. https://doi.org/10.1016/j.intell.2018.03.003

Glover, V. (2014). Maternal depression, anxiety and stress during pregnancy and child outcome; what needs to be done. *Best Pract Res Clin Obstet Gynaecol, 28*(1), 25-35. doi:10.1016/j.bpobgyn.2013.08.017

Goisis, A., Schneider, D. C., & Myrskylä, M. (2017). The reversing association between advanced maternal age and child cognitive ability: Evidence from three UK birth cohorts. *International Journal of Epidemiology*, *46*(3), 850–859. https://doi.org/10.1093/ije/dyw354

Goodman S. H. (2007). Depression in mothers. *Annual review of clinical psychology, 3*, 107–135. https://doi.org/10.1146/annurev.clinpsy.3.022806.091401

Goodman, S. H., & Gotlib, I. H. (1999). Risk for Psychopathology in the Children of Depressed Mothers : A Developmental Model for Understanding Mechanisms of Transmission. *Psychological Review, 106*(3), 458–490.

Grace, S. L., Evindar, A., & Stewart, D. E. (2003). Review article The effect of postpartum depression on child cognitive development and behavior : A review and critical analysis of the literature. *Archives of Womens Mental Health, 6,* 263–274. https://doi.org/10.1007/s00737-003-0024-6

Grant, K. A., McMahon, C., Reilly, N., & Austin, M. P. (2010). Maternal sensitivity moderates the impact of prenatal anxiety disorder on infant responses to the still-face procedure. *Infant Behavior and Development*, *33*(4), 453–462. https://doi.org/10.1016/j.infbeh.2010.05.001

Graignic-Philippe, R., Dayan, J., Chokron, S., Jacquet, A. Y., & Tordjman, S. (2014). Effects of prenatal stress on fetal and child development: a critical literature review. *Neurosci Biobehav Rev, 43*, 137-162. doi:10.1016/j.neubiorev.2014.03.022

Griffiths, R. (1954). The abilities of babies. London: University of London Press

Grol-Prokopczyk, H. (2018). In pursuit of anchoring vignettes that work: Evaluating generality versus specificity in vignette texts. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *73*(1), 54–63. https://doi.org/10.1093/geronb/gbx048

Grol-Prokopczyk, H., Freese, J., & Hauser, R. M. (2011). Using anchoring vignettes to assess group differences in general self-rated health. Journal of health and social behavior, 52(2), 246–261. https://doi.org/10.1177/0022146510396713.

Grol-Prokopczyk, H., Verdes-Tennant, E., McEniry, M., & Ispány, M. (2015). Promises and Pitfalls of Anchoring Vignettes in Health Survey Research. *Demography*, *52*(5), 1703–

1728. https://doi.org/10.1007/s13524-015-0422-1

Guerra, N., Williamson, A., & Lucas-Molina, B. (2012). Normal development: Infancy, childhood, and adolescence. Retrieved from: https://iacapap.org/content/uploads/A.2.-DEVELOPMENT-072012.pdf.

Gueron-Sela, N., Camerota, M., Willoughby, M. T., Vernon-Feagans, L., Cox, M. J., Greenberg, M. T., … Mills-Koonce, W. R. (2018). Maternal depressive symptoms, mother-child interactions, and children's executive function. *Developmental Psychology*, *54*(1), 71–82. https://doi.org/10.1037/dev0000389

Gur, R. C., Richard, J., Calkins, M. E., Chiavacci, R., Hansen, J. A., Bilker, W. B., … Gur, R. E. (2012). Age group and sex differences in performance on a computerized neurocognitive battery in children age 8-21. *Neuropsychology, 26*(2), 251–265. https://doi.org/10.1037/a0026712

Hamadani, J. D., Tofail, F., Hilaly, A., Mehrin, F., Shiraji, S., Banu, S., & Huda, S. N. (2012). Association of postpartum maternal morbidities with children's mental, psychomotor and language development in rural Bangladesh. *J Health Popul Nutr, 30*(2), 193-204. doi:10.3329/jhpn.v30i2.11313

Hanandita, W., & Tampubolon, G. (2016). Does reporting behaviour bias the measurement of social inequalities in self-rated health in Indonesia? An anchoring vignette analysis. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 25*(5), 1137–1149. https://doi.org/10.1007/s11136-015-1152-y

Hanley, G. E., Brain, U., & Oberlander, T. F. (2013). Infant developmental outcomes following prenatal exposure to antidepressants, and maternal depressed mood and positive affect. *Early Human Development*, *89*(8), 519–524. https://doi.org/10.1016/j.earlhumdev.2012.12.012

Hannon, P. (2003). Developmental neuroscience : implications for early childhood intervention and education. *Current Paediatrics, 13,* 58–63. https://doi.org/10.1054/cupe.2003.0410

Harzing, A. W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243–266. https://doi.org/10.1177/1470595806066332

Hay, D. F., Pawlby, S., Sharp, D., Asten, P., Mills, A., & Kumar, R. (2001). Intellectual problems shown by 11-year-old children whose mothers had postnatal depression. Journal of Child Psychology and Psychiatry and Allied Disciplines, 42(7), 871–889. https://doi.org/10.1111/1469-7610.00784

He, J., & van de Vijver, F. (2012). Bias and Equivalence in Cross-Cultural Research. *Online Readings in Psychology and Culture*, *2*(2), 1–19. https://doi.org/10.9707/2307-0919.1111

He, J., Van de Vijver, F. J. R., Fetvadjiev, V. H., de Carmen Dominguez Espinosa, A., Adams, B., Alonso-Arbiol, I., … Zhang, R. (2017). On Enhancing the Cross-Cultural Comparability of Likert-Scale Personality and Value Measures: A Comparison of Common Procedures. *European Journal of Personality*, *31*(6), 642–657. https://doi.org/10.1002/per.2132

Heiland, F. (2009). Does the birth order affect the cognitive development of a child?

Applied Economics, *41*(14), 1799–1818. https://doi.org/10.1080/00036840601083220

Heng, J., Quan, J., Sim, L. W., Sanmugam, S., Broekman, B., Bureau, J. F., … Rifkin-Graboi, A. (2018). The role of ethnicity and socioeconomic status in Southeast Asian mothers' parenting sensitivity. *Attachment and Human Development*, *20*(1), 24–42. https://doi.org/10.1080/14616734.2017.1365912

Hentges, F., Hill, J., Bishop, D. V. M., Goodacre, T., Moss, T., & Murray, L. (2011). The effect of cleft lip on cognitive development in school-aged children: A paradigm for examining sensitive period effects. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(6), 704–712. https://doi.org/10.1111/j.1469-7610.2011.02375.x

Herba, C. M., Glover, V., Ramchandani, P. G., & Rondon, M. B. (2016). Maternal depression and mental health in early childhood: an examination of underlying mechanisms in low-income and middle-income countries. *Lancet Psychiatry, 3*(10), 983-992. doi:10.1016/S2215-0366(16)30148-1

Hinz, A. (2017). Using Anchoring Vignettes in the Evaluation of Breast Cancer Survivors' Quality of Life. *Breast Care*, *12*(1), 34–38. https://doi.org/10.1159/000455002

Hoffman, C., Dunn, D. M., & Njoroge, W. F. M. (2017). Impact of Postpartum Mental Illness Upon Infant Development. *Current Psychiatry Reports, 19*(100)*,* 1–6.

Hopkins, D. J., & King, G. (2010). Improving anchoring vignettes designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*, *74*(2), 201–222. https://doi.org/10.1093/poq/nfq011

Howard, L. M., Molyneaux, E., Dennis, C. L., Rochat, T., Stein, A., & Milgrom, J. (2014). Non-psychotic mental disorders in the perinatal period. *The Lancet*, *384*(9956), 1775–1788. https://doi.org/10.1016/S0140-6736(14)61276-9

Husain, N., Cruickshank, J. K., Tomenson, B., Khan, S., & Rahman, A. (2012). Maternal depression and infant growth and development in British Pakistani women: a cohort study. *BMJ Open, 2*(2), e000523. doi:10.1136/bmjopen-2011-000523

Ibanez, G., Bernard, J. Y., Rondet, C., Peyre, H., Forhan, A., Kaminski, M., . . . Group, E. M.-C. C. S. (2015). Effects of Antenatal Maternal Depression and Anxiety on Children's Early Cognitive Development: A Prospective Cohort Study. *PLoS ONE, 10*(8), e0135849. doi:10.1371/journal.pone.0135849

Jensen, S. K. G., Dumontheil, I., & Barker, E. D. (2014). Developmental inter-relations between early maternal depression, contextual risks, and interpersonal stress, and their effect on later child cognitive functioning. *Depression and Anxiety*, *31*(7), 599–607. https://doi.org/10.1002/da.22147

Johnson J., Weissman M. M., & Klerman G. L. (1992) Service Utilization and Social Morbidity Associated With Depressive Symptoms in the Community. *JAMA, 267*(11), 1478–1483. doi:10.1001/jama.1992.03480110054033

Kang, E., & Grol-Prokopczyk, H. (2020). Comparing South Korean and US self-rated health using anchoring vignettes. *Quality of Life Research*. https://doi.org/10.1007/s11136-020-02599-y

Kaplan, P. S., Danko, C. M., Everhart, K. D., Diaz, A., Asherin, R. M., Vogeli, J. A. M., & Fekri, S. M. (2014). Maternal depression and expressive communication in one-year-old infants. *Infant Behavior and Development*, *37*(3), 398–405. https://doi.org/10.1016/j.infbeh.2014.05.008

Kapteyn, A., Smith, J. P., & Van Soest, A. (2007). Vignettes and self-reports of work disability in the United States and the Netherlands. *American Economic Review*, *97*(1), 461–473. https://doi.org/10.1257/aer.97.1.461

Kawai, E., Takagai, S., Takei, N., Itoh, H., Kanayama, N., & Tsuchiya, K. J. (2017). Maternal postpartum depressive symptoms predict delay in non-verbal communication in 14-month-old infants. *Infant Behavior and Development*, *46*, 33–45. https://doi.org/10.1016/j.infbeh.2016.11.006

Keim, S. A., Daniels, J. L., Dole, N., Herring, A. H., Siega-Riz, A. M., & Scheidt, P. C. (2011). A prospective study of maternal anxiety, perceived stress, and depressive symptoms in relation to infant cognitive development. *Early Human Development*, *87*(5), 373–380. https://doi.org/10.1016/j.earlhumdev.2011.02.004

Kiernan, K. E., & Huerta, M. C. (2008). Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood. *British Journal of Sociology*, *59*(4), 783–806. https://doi.org/10.1111/j.1468-4446.2008.00219.x

Kihlstrom, J. F. (2018). Unconscious Cognition☆. In Reference Module in Neuroscience and Biobehavioral Psychology: Elsevier.

King, G. (2009). Anhoring Vignettes: Frequently Asked Questions. Retrieved from: https://gking.harvard.edu/anchoring-vignettes-faqs

King, G., Murray C. J., Salomon J. A., and Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review 98*(1):191-207. http://nrs.harvard.edu/urn-3:HUL.InstRepos:3965182

King, G., & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis*, *15*(1), 46–66. https://doi.org/10.1093/pan/mpl011

Kingston, D., Mcdonald, S., Austin, M., & Tough, S. (2015). Association between Prenatal and Postnatal Psychological Distress and Toddler Cognitive Development : A Systematic Review. *PLoS ONE, 10*(5), 1–16. https://doi.org/10.1371/journal.pone.0126929

Kingston, D., Tough, S., & Whitfield, H. (2012). Prenatal and Postpartum Maternal Psychological Distress and Infant Development : A Systematic Review. *Child Psychiatry Hum Dev, 43,* 683–714. https://doi.org/10.1007/s10578-012-0291-4

Knott, R. J., Lorgelly, P. K., Black, N., & Hollingsworth, B. (2017a). Differential item functioning in quality of life measurement: An analysis using anchoring vignettes. *Social Science and Medicine*, *190*, 247–255. https://doi.org/10.1016/j.socscimed.2017.08.033

Knott, R. J., Black, N., Hollingsworth, B., & Lorgelly, P. K. (2017b). Response to comment by robone: Practical advice for the implementation of anchoring vignettes. *Health Economics (United Kingdom)*, *26*(3), 398–400. https://doi.org/10.1002/hec.3433

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *PNAS*, *103*(27), 10155–10162. https://doi.org/10.1073/pnas.0600888103

Koutra, K., Chatzi, L., Bagkeris, M., Vassilaki, M., Bitsios, P., & Kogevinas, M. (2013). Antenatal and postnatal maternal mental health as determinants of infant

neurodevelopment at 18 months of age in a mother-child cohort (Rhea Study) in Crete, Greece. *Social Psychiatry and Psychiatric Epidemiology*, *48*(8), 1335–1345. https://doi.org/10.1007/s00127-012-0636-0

Kovas, Y., Hayiou-Thomas, M. E., Oliver, B., Dale, P. S., Bishop, D. V., & Plomin, R. (2005). Genetic influences in different aspects of language development: the etiology of language skills in 4.5-year-old twins. Child Dev, 76(3), 632-651. doi:10.1111/j.1467-8624.2005.00868.x

Kowal, P., Chatterji, S., Naidoo, N., Biritwum, R., Fan, W., Ridaura, R. L., … Newell, M. L. (2012). Data resource profile: The world health organization study on global ageing and adult health (SAGE). *International Journal of Epidemiology*, *41*(6), 1639–1649. https://doi.org/10.1093/ije/dys210

Kristensen, P., & Bjerkedal, T. (2007). Explaining the relation between birth order and intelligence. *Science*, *316*(5832), 1717. https://doi.org/10.1126/science.1141493

Kristensen, N., & Johansson, E. (2008). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics*, *15*(1), 96–117. https://doi.org/10.1016/j.labeco.2006.11.001

Kurstjens, S., & Wolke, D. (2001). Effects of maternal depression on cognitive development of children over the first 7 years of life. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(5), 623–636. https://doi.org/10.1017/S0021963001007296

Lakens D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., … Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, *12*(5), 224–231. https://doi.org/10.1016/S0924-9338(97)83296-8

Lee, E. J. (2013). Differential susceptibility to the effects of child temperament on maternal warmth and responsiveness. *Journal of Genetic Psychology*, *174*(4), 429–449. https://doi.org/10.1080/00221325.2012.699008

Leijon, I., Ingemansson, F., Nelson, N., Wadsby, M., & Samuelsson, S. (2016). Reading deficits in very low birthweight children are associated with vocabulary and attention issues at the age of seven. Acta Paediatr, 105(1), 60-68. doi:10.1111/apa.13094

Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic Status and Child Development: A Meta-Analysis. *Journal of Emotional and Behavioral Disorders*, *21*(3), 211–224. https://doi.org/10.1177/1063426611421007

Lin, Y., Xu, J., Huang, J., Jia, Y., Zhang, J., Yan, C., & Zhang, J. (2017). Effects of prenatal and postnatal maternal emotional stress on toddlers' cognitive and temperamental development. *J Affect Disord, 207*, 9-17. doi:10.1016/j.jad.2016.09.010

Lindsay, K. L., Buss, C., Wadhwa, P. D., & Entringer, S. (2019). The Interplay Between Nutrition and Stress in Pregnancy: Implications for Fetal Programming of Brain Development. Biol Psychiatry, 85(2), 135-149. doi:10.1016/j.biopsych.2018.06.021

Liu, Y., Kaaya, S., Chai, J., Mccoy, D. C., Surkan, P. J., & Black, M. M., … Smith-Fawzi, M. C.

(2018). Maternal depressive symptoms and early childhood cognitive development : a meta-analysis. *Psychological Medicine, 47,* 680–689. https://doi.org/10.1017/S003329171600283X

Lovejoy, M. C., Graczyk, P. A., O'Hare, E., & Neuman, G. (2000). Maternal depression and parenting behavior. *Clinical Psychology Review*, *20*(5), 561–592. https://doi.org/10.1016/s0272-7358(98)00100-7

Lyons-Ruth, K., Zoll, D., Connell, D., & Grunebaum, H. U. (1986). The depressed mother and her one-year-old infant: environment, interaction, attachment, and infant development. *New directions for child development, 34*, 61–82. https://doi.org/10.1002/cd.232198634077

Magni, G., & Bilotta, F. (2016). Chapter 41 - Postoperative Cognitive Dysfunction. In H. Prabhakar (Ed.), Complications in Neuroanesthesia (pp. 411-427). San Diego: Academic Press.

Marksteiner, T., Kuger, S., & Klieme, E. (2018). The potential of anchoring vignettes to increase intercultural comparability of non-cognitive factors. *Assessment in Education: Principles, Policy and Practice*, *26*(4), 516–536. https://doi.org/10.1080/0969594X.2018.1514367

Marshall, R. E., Milligan-Saville, J., Petrie, K., Bryant, R. A., Mitchell, P. B., & Harvey, S. B. (2021). Mental health screening amongst police officers: factors associated with under-reporting of symptoms. BMC Psychiatry, 21(1), 135. doi:10.1186/s12888-021-03125-1

Martin, C. R., & Redshaw, M. (2018). Establishing a coherent and replicable measurement model of the Edinburgh Postnatal Depression Scale. *Psychiatry research, 264*, 182–191. https://doi.org/10.1016/j.psychres.2018.03.062

Menon, B., Cherkil, S., Aswathy, S., Unnikrishnan, A., & Rajani, G. (2012). The process and challenges in the translation of World Health Organization Quality of Life (WHOQOL-BREF) to a regional language; Malayalam. *Indian Journal of Psychological Medicine*, *34*(2), 149–152. https://doi.org/10.4103/0253-7176.101783

Mesman, J., & Emmen, R. A. G. (2013). Mary Ainsworth's legacy: A systematic review of observational instruments measuring parental sensitivity. *Attachment and Human Development*, *15*(5–6), 485–506. https://doi.org/10.1080/14616734.2013.820900

Mesman, J., Minter, T., & Angnged, A. (2016). Received sensitivity: adapting Ainsworth's scale to capture sensitivity in a multiple-caregiver context. *Attachment & Human Development, 18*(2), 101-114. doi:10.1080/14616734.2015.1133681

Mesman, J., Minter, T., Angnged, A., Cissé, I. A. H., Salali, G. D., & Migliano, A. B. (2018). Universality Without Uniformity: A Culturally Inclusive Approach to Sensitive Responsiveness in Infant Caregiving. *Child Development*, *89*(3), 837–850. https://doi.org/10.1111/cdev.12795

Mesman, J., van Ijzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2012). Unequal in Opportunity, Equal in Process: Parental Sensitivity Promotes Positive Child Development in Ethnic Minority Families. *Child Development Perspectives*, *6*(3), 239–250. https://doi.org/10.1111/j.1750-8606.2011.00223.x

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: applications in cross-cultural research. *International Journal of Psychological*

Research, *3*(1), 111–130. https://doi.org/10.21500/20112084.857

Milgrom, J., Westley, D. T., & Gemmill, A. W. (2004). The mediating role of maternal responsiveness in some longer term effects of postnatal depression on infant development. *Infant Behavior and Development*, *27*(4), 443–454. https://doi.org/10.1016/j.infbeh.2004.03.003

Mills-Koonce, W. R., Gariepy, J. L., Sutton, K., & Cox, M. J. (2008). Changes in maternal sensitivity across the first three years: Are mothers from different attachment dyads differentially influenced by depressive symptomatology? *Attachment and Human Development*, *10*(3), 299–317. https://doi.org/10.1080/14616730802113612

Milner, J. (2016). Mental health in China and India: a growing storm. *The Lancet Psychiatry*, *3*(9), 793–794. https://doi.org/10.1016/S2215-0366(16)30173-0

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Mojtabai, R. (2016). Depressed Mood in Middle-Aged and Older Adults in Europe and the United States: A Comparative Study Using Anchoring Vignettes. *Journal of Aging and Health*, *28*(1), 95–117. https://doi.org/10.1177/0898264315585506

Molfese, V. J., DiLalla, L. F., & Lovelace, L. (1996). Perinatal, Home Environment, and Infant Measures as Successful Predictors of Preschool Cognitive and Verbal Abilities. *International Journal of Behavioral Development*, *19*(1), 101–119. https://doi.org/10.1177/016502549601900109

Molina, T. (2016). Reporting Heterogeneity and Health Disparities Across Gender and Education Levels: Evidence From Four Countries. *Demography*, *53*(2), 295–323. https://doi.org/10.1007/s13524-016-0456-z

Molina, T. (2017). Adjusting for heterogeneous response thresholds in cross-country comparisons of self-reported health. *Journal of the Economics of Ageing*, *10*, 1–20. https://doi.org/10.1016/j.jeoa.2017.08.002

Morris, K. A. (2018). Measurement equivalence: A glossary for comparative population health research. *Journal of Epidemiology and Community Health*, *72*(7), 559–563. https://doi.org/10.1136/jech-2017-209962

Mõttus, R., Allik, J., Realo, A., Rossier, J., Zecca, G., Ah-Kion, J., … Johnson, W. (2012). The Effect of Response Style on Self-Reported Conscientiousness Across 20 Countries. *Personality and Social Psychology Bulletin*, *38*(11), 1423–1436. https://doi.org/10.1177/0146167212451275

Mullen, E. M. (1995). *Mullen Scales of Early Learning* (AGS ed.). Circle Pines, MN: American Guidance Service Inc.

Munoz-Rocha, T. V., Tamayo, Y. O. M., Romero, M., Pantic, I., Schnaas, L., Bellinger, D., . . . Tellez-Rojo, M. M. (2018). Prenatal co-exposure to manganese and depression and 24-months neurodevelopment. *NeuroToxicology, 64*, 134-141. doi:10.1016/j.neuro.2017.07.007

Murray, L. (1992). The Impact of Postnatal Depression on Infant Development. *Journal of Child Psychology and Psychiatry*, *33*(3), 543–561. https://doi.org/10.1111/j.1469-7610.1992.tb00890.x

Murray, L. (1997). Postpartum Depression and Child Development. *Psychological Medicine*, *27*(2), 253–260. https://doi.org/10.1017/S0033291796004564

Murray, L., Arteche, A., Fearon, P., Halligan, S., Croudace, T., & Cooper, P. (2010). The effects of maternal postnatal depression and child sex on academic performance at age 16 years: A developmental approach. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *51*(10), 1150–1159. https://doi.org/10.1111/j.1469-7610.2010.02259.x

Murray, L., Cooper, P., Arteche, A., Stein, A., & Tomlinson, M. (2016). Randomized controlled trial of a home-visiting intervention on infant cognitive development in peri-urban South Africa. *Dev Med Child Neurol, 58*(3), 270-276. doi:10.1111/dmcn.12873

Murray, L., Fearon, P., & Cooper, P. (2015). Postnatal Depression, Mother–Infant Interactions, and Child Development Prospects for Screening and Treatment Effects of Postnatal Depression on Mother–Child Relationships Early infancy. In J. Milgrom & A. Gemmill (Eds.), *Identifying Perinatal Depression and Anxiety: Evidence-Based Practice in Screening, Psychosocial Assessment and Management*: John Wiley & Sons.

Murray, L., Fiori-Cowley, A., Hooper, R., & Cooper, P. (1996). The Impact of Postnatal Depression and Associated Adversity on Early Mother-Infant Interactions and Later Infant Outcome. *Child Development*, *67*(5), 2512–2526. https://doi.org/10.1111/j.1467-8624.1996.tb01871.x

Murray, C. J. L., Health, W., & Salomon, J. a. (2004). of Measurement in Survey Research. *American Political Science Review*, *98*(1), 191–207. https://doi.org/10.1017/S0003055403000881

Murray, L., Hentges, F., Hill, J., Karpf, J., Mistry, B., Kreutz, M., … Green, R. (2008). The effect of cleft lip and palate, and the timing of lip repair on mother-infant interactions and infant development. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(2), 115–123. https://doi.org/10.1111/j.1469-7610.2007.01833.x

Murray, L., Kempton, C., Woolgar, M., & Hooper, R. (1993). Depressed Mothers' Speech to Their Infants and its Relation to Infant Gender and Cognitive Development. *Journal of Child Psychology and Psychiatry*, *34*(7), 1083–1101. https://doi.org/10.1111/j.1469-7610.1993.tb01775.x

Murray, C., Özaltin, E., Tandon, A., Salomon, J., Sadana, R., & Chatterji, S. (2003). Empirical evaluation of the anchoring vignettes approach in health surveys. In *Health Systems Performance Assessment: Debates, Methods and Empiricism*, Murray CJL, Evans DB (eds). World Health Organization: Geneva.

Najman, J. M., Aird, R., Bor, W., O'Callaghan, M., Williams, G. M., & Shuttlewood, G. J. (2004). The generational transmission of socioeconomic inequalities in child cognitive development and emotional health. *Social Science and Medicine*, *58*(6), 1147–1158. https://doi.org/10.1016/S0277-9536(03)00286-7

National Institute for Health and Care Excellence. Antenatal and postnatalmental health: Clinical management and service guidance. NICEguidelines [CG45]. February, 2007. Available from: http://guidance.nice.org.uk/CG45/Guidance.

Nelson, C. A., 3rd, Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A. T., & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: the Bucharest Early

Intervention Project. *Science, 318*(5858), 1937–1940. https://doi.org/10.1126/science.1143921

NICHD Early Child Care Research Network. (1999). Chronicity of maternal depressive symptoms, maternal sensitivity, and child functioning at 36 months. *Developmental Psychology, 35*(5), 1297–1310. doi:10.1037/0012-1649.35.5.1297

Noble, M., Wright, G., Dibben, C., Smith, G., McLennan, D., Anttila, C., et al. (2004). The English Indices of Deprivation 2004 (revised). Report to the Office of the Deputy Prime Minister. London: Neighbourhood Renewal Unit.

Norhayati, M. N., Nik Hazlina, N. H., Asrenee, A. R., & Wan Emilin, W. M. A. (2015). Magnitude and risk factors for postpartum symptoms: A literature review. *Journal of Affective Disorders*, *175*, 34–52. https://doi.org/10.1016/j.jad.2014.12.041

O'Hara, M. W., & Swain, A. M. (1996). Rates and risk of postpartum depression—a meta-analysis. *International Review of Psychiatry, 8*(1), 37-54. doi:10.3109/09540269609037816Johnson J,

O'Leary, N., Jairaj, C., Molloy, E. J., McAuliffe, F. M., Nixon, E., & O'Keane, V. (2019). Antenatal depression and the impact on infant cognitive, language and motor development at six and twelve months postpartum. *Early Hum Dev, 134*(May), 41-46. doi:10.1016/j.earlhumdev.2019.05.021

Orchinik, L. J., Taylor, H. G., Espy, K. A., Minich, N., Klein, N., Sheffield, T., & Hack, M. (2011). Cognitive outcomes for extremely preterm/extremely low birth weight children in kindergarten. J Int Neuropsychol Soc, 17(6), 1067-1079. doi:10.1017/s135561771100107x

Osborne, S., Biaggi, A., Chua, T. E., Du Preez, A., Hazelgrove, K., Nikkheslat, N., … Pariante, C. M. (2018). Antenatal depression programs cortisol stress reactivity in offspring through increased maternal inflammation and cortisol in pregnancy: The Psychiatry Research and Motherhood – Depression (PRAM-D) Study. *Psychoneuroendocrinology*, *98*, 211–221. https://doi.org/10.1016/j.psyneuen.2018.06.017

Otake, Y., Nakajima, S., Uno, A., Kato, S., Sasaki, S., Yoshioka, E., . . . Kishi, R. (2014). Association between maternal antenatal depression and infant development: a hospital-based prospective cohort study. *Environ Health Prev Med, 19*(1), 30-45. doi:10.1007/s12199-013-0353-7

Paccagnella, O., (2013). Modelling individual heterogeneity in ordered choice models: Anchoring Vignettes and the Chopit Model. *Quaderni Di Statistica*, *15*(15), 69–94.

Pan, Y., & Fond, M. (2014). Evaluating multilingual questionnaires: A sociolinguistic perspective. *Survey Research Methods*, *8*(3), 181–194. https://doi.org/10.18148/srm/2014.v8i3.5483

Park, M., Brain, U., Grunau, R. E., Diamond, A., & Oberlander, T. F. (2018). Maternal depression trajectories from pregnancy to 3 years postpartum are associated with children's behavior and executive functions at 3 and 6 years. *Archives of Women's Mental Health*, *21*(3), 353–363. https://doi.org/10.1007/s00737-017-0803-0

Parker, G., Cheah, Y. C., & Roy, K. (2001). Do The Chinese Somatize Depression. *Social Psychiatry and Psychiatric Epidemiology*, *36*, 287–293.

Parsons, C. E., Young, K. S., Rochat, T. J., Kringelbach, M. L., & Stein, A. (2012). Postnatal depression and its effects on child development: a review of evidence from low-

and middle-income countries. *Br Med Bull, 101*(May), 57-79. doi:10.1093/bmb/ldr047

Patel, V., DeSouza, N., & Rodrigues, M. (2003). Postnatal depression and infant growth and development in low income countries: a cohort study from Goa, India. *Arch Dis Child, 88*(1), 34-37. doi:10.1136/adc.88.1.34

Patni, B. (2012). Developmental assessment scales for Indian infants (DASII). Indian Journal of Practical Pediatrics. 14. 409-412.

Pearson, R. M., Melotti, R., Heron, J., Joinson, C., Stein, A., Ramchandani, P. G., & Evans, J. (2012). Disruption to the development of maternal responsiveness? The impact of prenatal depression on mother-infant interactions. *Infant Behavior and Development*, *35*(4), 613–626. https://doi.org/10.1016/j.infbeh.2012.07.020

Perra, O., Phillips, R., Fyfield, R., Waters, C., & Hay, D. F. (2015). Does mothers' postnatal depression influence the development of imitation? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(11), 1231–1238. https://doi.org/10.1111/jcpp.12413

Petterson, S. M., & Albers, A. B. (2001). Effects of poverty and maternal depression on early child development. *Child Development*, *72*(6), 1794–1813. https://doi.org/10.1111/1467-8624.00379

Piteo, A. M., Yelland, L. N., & Makrides, M. (2012). Does maternal depression predict developmental outcome in 18month old infants? *Early Human Development*, *88*(8), 651–655. https://doi.org/10.1016/j.earlhumdev.2012.01.013

Poksinska, B. B., & Cronemyr, P. (2017). Measuring quality in elderly care: possibilities and limitations of the vignette method. *Total Quality Management and Business Excellence*, *28*(9–10), 1194–1207. https://doi.org/10.1080/14783363.2017.1303875

Pradhan, S. C., Singh, M. M., Singh, R. A., Das, J., Ram, D., Patil, B., Jain, A. K., & Thomas, J. K. (2001). First care givers of mentally ill patients: a multicenter study. *Indian journal of medical sciences, 55*(4), 203–208.

Putnick, Diane, L., & Bornstein, Mark, H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.Measurement

Quevedo, L. A., Silva, R. A., Godoy, R., Jansen, K., Matos, M. B., Tavares Pinheiro, K. A., & Pinheiro, R. T. (2012). The impact of maternal post-partum depression on the language development of children at 12 months. *Child Care Health Dev, 38*(3), 420-424. doi:10.1111/j.1365-2214.2011.01251.x

Radloff, L. S. (1977). The CES-D scale: A self report depression scale for research in the general population. Applied Psychological Measurements, 1, 385-401.

Rahman, A., & Creed, F. (2007). Outcome of prenatal depression and risk factors associated with persistence in the first postnatal year: Prospective study from Rawalpindi, Pakistan. *Journal of Affective Disorders*, *100*(1–3), 115–121. https://doi.org/10.1016/j.jad.2006.10.004

Rice, N., Robone, S., & Smith, P. (2011). Analysis of the validity of the vignette approach to correct for heterogeneity in reporting health system responsiveness. *European Journal of Health Economics*, *12*(2), 141–162. https://doi.org/10.1007/s10198-010-

0235-5

Richter, L. M., Daelmans, B., Lombardi, J., Heymann, J., Boo, F. L., Behrman, J. R., … Darmstadt, G. L. (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *The Lancet*, *389*(10064), 103–118. https://doi.org/10.1016/S0140-6736(16)31698-1

Rodgers, J. L. (2014). Are birth order effects on intelligence really Flynn Effectsα Reinterpreting Belmont and Marolla 40 years later. *Intelligence*, *42*(1), 128–133. https://doi.org/10.1016/j.intell.2013.08.004

Rodgers, J. L., Cleveland, H. H., Van Den Oord, E., & Rowe, D. C. (2000). Resolving the debate over birth order, family size, and intelligence. *American Psychologist*, *55*(6), 599–612. https://doi.org/10.1037/0003-066X.55.6.599

Rogler, L. H. (1999). Methodological sources of cultural insensitivity in mental health research. *American Psychologist*, *54*(6), 424–433. https://doi.org/10.1037/0003-066X.54.6.424

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of Infant Cognition for Executive Functions at Age 11. *Psychological Science*, *23*(11), 1345–1355. https://doi.org/10.1177/0956797612444902

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2012). Information processing from infancy to 11years: Continuities and prediction of IQ. *Intelligence*, *40*(5), 445–457. https://doi.org/10.1016/j.intell.2012.05.007

Rotheram-Fuller, E. J., Tomlinson, M., Scheffler, A., Weichle, T. W., Hayati Rezvan, P., Comulada, W. S., & Rotheram-Borus, M. J. (2018). Maternal patterns of antenatal and postnatal depressed mood and the impact on child health at 3-years postpartum. *J Consult Clin Psychol, 86*(3), 218-230. doi:10.1037/ccp0000281

Rossouw, L., Bago d'Uva, T., & van Doorslaer, E. (2018). Poor Health Reporting? Using Anchoring Vignettes to Uncover Health Disparities by Wealth and Race. *Demography*, *55*(5), 1935–1956. https://doi.org/10.1007/s13524-018-0709-0

Sagar, R., Dandona, R., Gururaj, G., Dhaliwal, R. S., Singh, A., Ferrari, A., … Dandona, L. (2020). The burden of mental disorders across the states of India: the Global Burden of Disease Study 1990–2017. *The Lancet Psychiatry*, *7*(2), 148–161. https://doi.org/10.1016/S2215-0366(19)30475-4

Saha, S., Barnett, A. G., Foldi, C., Burne, T. H., Eyles, D. W., Buka, S. L., & McGrath, J. J. (2009). Advanced paternal age is associated with impaired neurocognitive outcomes during infancy and childhood. *PLoS Medicine*, *6*(3), 0303–0311. https://doi.org/10.1371/journal.pmed.1000040

Salomon, J. A., Tandon, A., & Murray, C. J. L. (2004). Comparability of self rated health: Cross sectional multi-country survey using anchoring vignettes. *British Medical Journal*, *328*(7434), 258–261. https://doi.org/10.1136/bmj.37963.691632.44

Sameroff, A. (2010). A unified theory of development: A dialectic integration of nature and nurture. *Child Development, 81*(1), 6–22. https://doi.org/10.1111/j.1467-8624.2009.01378.x

Schechter, J. C., Brennan, P. A., Smith, A. K., Stowe, Z. N., Newport, D. J., & Johnson, K. C. (2017). Maternal Prenatal Psychological Distress and Preschool Cognitive

Functioning: the Protective Role of Positive Parental Engagement. *J Abnorm Child Psychol, 45*(2), 249-260. doi:10.1007/s10802-016-0161-9

Sen, A. (2002). Health: perception versus observation. *British Medical Journal*, *324*(7342), 860–861. https://doi.org/10.1136/bmj.324.7342.860

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Sharma, D. C. (2014). India's new policy aims to close gaps in mental health care. *The Lancet*, *384*(9954), 1564. https://doi.org/10.1016/S0140-6736(14)61973-5

Sharp, D., Hay, D. F., Pawlby, S., Schmücker, G., Allen, H., & Kumar, R. (1995). The Impact of Postnatal Depression on Boys' Intellectual Development. *Journal of Child Psychology and Psychiatry*, *36*(8), 1315–1336. https://doi.org/10.1111/j.1469-7610.1995.tb01666.x

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. The Journal of Clinical Psychiatry, 59(Suppl 20), 22-33

Shorey, S., Chee, C. Y. I., Ng, E. D., Chan, Y. H., Tam, W. W. S., & Chong, Y. S. (2018). Prevalence and incidence of postpartum depression among healthy mothers: A systematic review and meta-analysis. *Journal of Psychiatric Research, 104*, 235-248. doi:https://doi.org/10.1016/j.jpsychires.2018.08.001

Skurtveit, S., Selmer, R., Roth, C., Hernandez-Diaz, S., & Handal, M. (2014). Prenatal exposure to antidepressants and language competence at age three: results from a large population-based pregnancy cohort in Norway. *BJOG, 121*(13), 1621-1631. doi:10.1111/1471-0528.12821

Smith-Nielsen, J., Tharner, A., Krogh, M. T., & Væver, M. S. (2016). Effects of maternal postpartum depression in a well-resourced sample: Early concurrent and long-term effects on infant cognitive, language, and motor development. *Scandinavian Journal of Psychology*, *57*(6), 571–583. https://doi.org/10.1111/sjop.12321

Stanley, C., Murray, L., & Stein, A. (2004). The effect of postnatal depression on mother-infant interaction, infant response to the Still-face perturbation, and performance on an Instrumental Learning task. *Development and Psychopathology*, *16*(1), 1–18. https://doi.org/10.1017/S0954579404040386

Stein, A., Craske, M. G., Lehtonen, A., Harvey, A., Savage-McGlynn, E., Davies, B., … Counsell, N. (2012). Maternal cognitions and mother-infant interaction in postnatal depression and generalized anxiety disorder. *Journal of Abnormal Psychology*, *121*(4), 795–809. https://doi.org/10.1037/a0026847

Stein, A., Malmberg, L. E., Sylva, K., Barnes, J., Leach, P., & Team, F. C. C. C. (2008). The influence of maternal depression, caregiving, and socioeconomic status in the post-natal year on children's language development. *Child: Care, Health and Development*, *34*(5), 603–612. https://doi.org/10.1111/j.1365-2214.2008.00837.x

Stein, A., Pearson, R. M., Goodman, S. H., Rapa, E., Rahman, A., McCallum, M., . . . Pariante, C. M. (2014). Effects of perinatal mental disorders on the fetus and child. *The Lancet, 384*(9956), 1800-1819. doi:10.1016/s0140-6736(14)61277-0

Su, Y., Willis, G., & Salomon, J. A. (2017). Improving Vignette Descriptions and Question Formats to Measure Distance Vision: Evidence from Cognitive Interviews among Students in China. *Field Methods*, *29*(3), 175–193. https://doi.org/10.1177/1525822X16680810

Surkan, P. J., Patel, S. A., & Rahman, A. (2016). Preventing infant and child morbidity and mortality due to maternal depression. *Best Practice and Research: Clinical Obstetrics and Gynaecology*, *36*, 156–168. https://doi.org/10.1016/j.bpobgyn.2016.05.007

Sutter-Dallay, A. L., Murray, L., Dequae-Merchadou, L., Glatigny-Dallay, E., Bourgeois, M. L., & Verdoux, H. (2011). A prospective longitudinal study of the impact of early postnatal vs. chronic maternal depressive symptoms on child development. *European Psychiatry*, *26*(8), 484–489. https://doi.org/10.1016/j.eurpsy.2010.05.004

Swingler, M. M., Perry, N. B., Calkins, S. D., & Bell, M. A. (2017). Maternal behavior predicts infant neurophysiological and behavioral attention processes in the first year. *Developmental Psychology*, *53*(1), 13–27. https://doi.org/10.1037/dev0000187

Taylor, L. (2005). Introducing Cognitive Development. London: Psychology Press, https://doi.org/10.4324/9780203504550

Tearne, J. E. (2015). Older maternal age and child behavioral and cognitive outcomes: A review of the literature. *Fertility and Sterility*, *103*(6), 1381–1391. https://doi.org/10.1016/j.fertnstert.2015.04.027

Teicher, M. H., & Samson, J. A. (2016). *Annual Research Review : Enduring neurobiological effects of childhood abuse and neglect*. *Journal of Child Psychology and Psychiatry, 57*(3), 241–266. https://doi.org/10.1111/jcpp.12507

Thippeswamy, H., Desai, G., & Chandra, P. (2018). Help-seeking patterns in women with postpartum severe mental illness: a report from southern India. *Archives of Women's Mental Health*, *21*(5), 573–578. https://doi.org/10.1007/s00737-018-0830-5

Thomas, S., Vigil, E., Thomas, T., Bellinger, D. C., Ramthal, A., Kurpad, A. V., . . . Srinivasan, K. (2020). Antenatal Depressive Symptoms and Neurodevelopment Outcomes in Children at 30 Months. A Study From South India. *Frontiers in Psychiatry, 11*, 1025. https://doi.org/10.3389/fpsyt.2020.486175

Tong, S., Baghurst, P., Vimpani, G., & McMichael, A. (2007). Socioeconomic Position, Maternal IQ, Home Environment, and Cognitive Development. *Journal of Pediatrics*, *151*(3). https://doi.org/10.1016/j.jpeds.2007.03.020

Topp, J., Heesen, C., Augustin, M., Andrees, V., & Blome, C. (2020). Challenges and lessons learned from using anchoring vignettes to explore quality of life response behavior. *Quality of Life Research*, *29*(8), 2149–2159. https://doi.org/10.1007/s11136-020-02488-4

Tout, K., De Haan, M., Campbell, E. K., & Gunnar, M. R. (1998). Social Behavior Correlates of Cortisol Activity in Child Care: Gender Differences and Time-of-Day Effects. *Child Development*, *69*(5), 1247–1262. https://doi.org/10.1111/j.1467-8624.1998.tb06209.x

Tran, T. D., Biggs, B. A., Tran, T., Simpson, J. A., Hanieh, S., Dwyer, T., & Fisher, J. (2013). Impact on infants' cognitive development of antenatal exposure to iron deficiency disorder and common mental disorders. *PLoS ONE, 8*(9), e74876. doi:10.1371/journal.pone.0074876

Tronick, E., & Reck, C. (2009). Infants of depressed mothers. *Harvard Review of Psychiatry*,

17(2), 147–156. https://doi.org/10.1080/10673220902899714

Tse, A. C., Rich-Edwards, J. W., Rifas-Shiman, S. L., Gillman, M. W., & Oken, E. (2010). Association of maternal prenatal depressive symptoms with child cognition at age 3 years. *Paediatr Perinat Epidemiol, 24*(3), 232-240. doi:10.1111/j.1365-3016.2010.01113.x

Underwood, L., Waldie, K., D'Souza, S., Peterson, E. R., & Morton, S. (2016). A review of longitudinal studies on antenatal and postnatal depression. *Archives of Women's Mental Health*, *19*(5), 711–720. https://doi.org/10.1007/s00737-016-0629-1

UNICEF. (2014). Early childhood development: A statistical snapshot. Building better brains and sustainable outcomes for children. *Statistical Snapshot*, 8. Retrieved from UNICEF/NYHQ2006-0162/Kamber

Upadhyay, R. P., Chowdhury, R., Salehi, A., Sarkar, K., Singh, S. K., Sinha, B., … Kumar, A. (2017). Postpartum depression in india: A systematic review and meta-analysis. *Bulletin of the World Health Organization*, *95*(10), 706–717. https://doi.org/10.2471/BLT.17.192237

Upadhyay, R. P., Naik, G., Choudhary, T. S., Chowdhury, R., Taneja, S., Bhandari, N., . . . Bhan, M. K. (2019). Cognitive and motor outcomes in children born low birth weight: a systematic review and meta-analysis of studies from South Asia. BMC Pediatrics, 19(1), 35. doi:10.1186/s12887-019-1408-8

Van den Bergh, B. R. H., van den Heuvel, M. I., Lahti, M., Braeken, M., de Rooij, S. R., Entringer, S., . . . Schwab, M. (2017). Prenatal developmental origins of behavior and mental health: The influence of maternal stress in pregnancy. *Neurosci Biobehav Rev*(November 2016). doi:10.1016/j.neubiorev.2017.07.003

Van De Vijver, F. J. R., & Poortinga, Y. H. (1997). van der Vijver (1997) Towards and integrated analysis of Bias in cross cultural assessment.pdf. *European Journal of Psychological Assessment*, *13*(1), 29–37.

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., & Smith, J. P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *174*(3), 575–595. https://doi.org/10.1111/j.1467-985X.2011.00694.x

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–69. https://doi.org/10.1177/109442810031002

Varma, D., Chandra, P. S., Thomas, T., & Carey, M. P. (2007). Intimate partner violence and sexual coercion among pregnant women in India: Relationship with depression and post-traumatic stress disorder. *Journal of Affective Disorders*, *102*(1–3), 227–235. https://doi.org/10.1016/j.jad.2006.09.026

Victora, C. G., Bahl, R., Barros, A. J., França, G. V., Horton, S., Krasevec, J., . . . Rollins, N. C. (2016). Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet, 387(10017), 475-490. doi:10.1016/s0140-6736(15)01024-7

Victora, C. G., Horta, B. L., Loret de Mola, C., Quevedo, L., Pinheiro, R. T., Gigante, D. P., . . . Barros, F. C. (2015). Association between breastfeeding and intelligence, educational attainment, and income at 30 years of age: a prospective birth cohort study from Brazil. Lancet Glob Health, 3(4), e199-205. doi:10.1016/s2214-109x(15)70002-1

Vonkova, H., Bendl, S., & Papajoanu, O. (2017). How Students Report Dishonest Behavior in School: Self-Assessment and Anchoring Vignettes. *Journal of Experimental Education*, *85*(1), 36–53. https://doi.org/10.1080/00220973.2015.1094438

Voňková, H., & Hullegie, P. (2011). Is the anchoring vignette method sensitive to the domain and choice of the vignette? *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *174*(3), 597–620. https://doi.org/10.1111/j.1467-985X.2011.00704.x

Vonkova, H., Zamarro, G., & Hitt, C. (2018). Cross-Country Heterogeneity in Students' Reporting Behavior: The Use of the Anchoring Vignette Method. *Journal of Educational Measurement*, *55*(1), 3–31. https://doi.org/10.1111/jedm.12161

Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: direct and moderated effects. *Child development, 75*(1), 296–312. https://doi.org/10.1111/j.1467-8624.2004.00670.x

Wachs, T. D., Black, M. M., & Engle, P. L. (2009). Maternal Depression : A Global Threat to Children ' s Health , Development , and Behavior and to Human Rights. *Child Development Perspectives, 3*(1), 51–59.

Walker, S. P., Wachs, T. D., Grantham-Mcgregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., … Richter, L. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, *378*(9799), 1325–1338. https://doi.org/10.1016/S0140-6736(11)60555-2

Weber, A., Darmstadt, G. L., & Rao, N. (2017). Gender disparities in child development in the east Asia-Pacific region: a cross-sectional, population-based, multicountry observational study. *The Lancet Child and Adolescent Health*, *1*(3), 213–224. https://doi.org/10.1016/S2352-4642(17)30073-1

Weinberg, M. K., Olson, K. L., Beeghly, M., & Tronick, E. Z. (2006). Making up is hard to do, especially for mothers with high levels of depressive symptoms and their infant sons. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(7), 670–683. https://doi.org/10.1111/j.1469-7610.2005.01545.x

Weiss, S., & Roberts, R. D. (2018). Using anchoring vignettes to adjust self-reported personality: A comparison between countries. *Frontiers in Psychology*, *9*(MAR), 1–17. https://doi.org/10.3389/fpsyg.2018.00325

Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses, http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp, 2013.

Whiffen, V. E., & Gotlib, I. H. (1989). Infants of Postpartum Depressed Mothers: Temperament and Cognitive Status. *Journal of Abnormal Psychology*, *98*(3), 274–279. https://doi.org/10.1037/0021-843x.98.3.274

Williamson, W. D., Wilson, G. S., Lifschitz, M. H., Thurber, S. A., 1990. Nonhandicapped very-low-birth-weight infants at one year of age: developmental profile. *Pediatrics*, 85, 405–410.

Windsor, J., Benigno, J. P., Wing, C. A., Carroll, P. J., Koga, S. F., Nelson, C. A., … Zeanah, C. H. (2011). Effect of foster care on young children's language learning. *Child Development*, *82*(4), 1040–1046. https://doi.org/10.1111/j.1467-8624.2011.01604.x

Woldarsky, V., Urzúa, C., Farkas, C., & Vallotton, C. D. (2019). Differences in Chilean and USA Mothers' Sensitivity Considering Child Gender and Temperament. *Journal of Child and Family Studies*, *28*(7), 1937–1947. https://doi.org/10.1007/s10826-019-01419-3

Woody, C. A., Ferrari, A. J., Siskind, D. J., Whiteford, H. A., & Harris, M. G. (2017). A systematic review and meta-regression of the prevalence and incidence of perinatal depression. *Journal of Affective Disorders*, *219*(May), 86–92. https://doi.org/10.1016/j.jad.2017.05.003

World Health Organisation (2008). mhGAP: Mental Health Gap Action Programme: Scaling Up Care for Mental, Neurological and Substance Use Disorders. Geneva. Available from: https://www.ncbi.nlm.nih.gov/books/NBK310851/

Zhang, S., Dang, R., Yang, N., Bai, Y., Wang, L., Abbey, C., & Rozelle, S. (2018). Effect of caregiver's mental health on early childhood development across different rural communities in china. *International Journal of Environmental Research and Public Health*, *15*(11), 1–25. https://doi.org/10.3390/ijerph15112341

Zung, W. Y. K. (1965). A self-rating depression scale. Archives of General Psychiatry, 12, 63-70.

Zung, W. Y. K. (1971). A rating instrument for anxiety disorders. Psychosomatics, 12, 371-379.

Appendix 1: Ethical approval letters

WCHADS (Phase 1-8 & Ammendment)

Cheshire North & West Research Ethics Committee
Cheshire West PCT
1829 Building
Countess of Chester Health Park
Liverpool Road
Chester
CH2 1HJ

Telephone: 01244 650 334
Facsimile: 01244 650 333

27 June 2006

Professor Jonathan Hill
Professor of Child and Developmental Psychiatry
University of Liverpool, Alder Hey Hospital
Mulberry House, Alder Hey Hospital
Eaton Road
L12 2AP

Dear Professor Hill

Full title of study: **The Wirral Child Health and Development Study**
REC reference number: **05/Q1506/107**

Thank you for your letter of 19 May 2006, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Vice-Chairman.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Application		09 January 2006
Investigator CV		
Protocol	1	09 January 2006
Covering Letter		09 January 2006
Summary/Synopsis	1	09 January 2006
Response to Request for Further Information		19 May 2006
Father Information Sheet, Study 1500 - Phases 1, 3, 5 & 7	2	01 May 2006
Study 300 Parent Information Sheet, one year - Phase 8	2	01 May 2006
Study 300 Parent Information Sheet, 6 months - Phase 6	2	01 May 2006

Study 300 Parent Information Sheet, Antenatal Phases 2 & 4	2	01 May 2006
Mother Information Sheet, Study 1500 - Phases 1, 3, 5, & 7	2	01 May 2006
Letter confirming funding - MRC		09 March 2005
Supporting letter from Mr Doyle, Wirral Hospitals NHS Trust		09 December 2005
Supporting letter from Ms Sheila Hillhouse, Birkenhead & Wallasey PCT		09 December 2005
Phase 8: Study 300 12 month mother and baby postnatal assessments	1	09 January 2006
GP Letter Study 1500	1	01 January 2006
GP Letter Study 300		01 January 2006
Parent Consent, Study 1500 - Phases 1, 3, 5 & 7	1	09 January 2006
Consent to contact a relative - Study 1500	1	09 January 2006
Parent Consent, Fathers, - Study 1500 - Phases 1, 3, 5 & 7	1	09 January 2006
Parent Consent - Study 300 Antenatal, perinatal - (Phases 2 & 4)	1	09 January 2006
Study 300 Parent Information Sheet 6 months (Phase 6)	1	09 January 2006
Parent Consent - Study 300, first birthday (Phase 8)	1	09 January 2006
Parent Consent - Study 300, DNA First Birthday (Phase 8)	1	09 January 2006
Phase 1: Study 1500 mother antenatal screen	1	09 January 2006
Phase 1: Study 1500 father antenatal screen	1	09 January 2006
Phase 2: Study 300 mother antenatal interview	1	09 January 2006
Phase 3: Study 1500 pregnancy/obstetric/birth outcomes	1	09 January 2006
Phase 4: Study 300 perinatal baby assessment	1	09 January 2006
Phase 5; Study 1500 6-8 week questionnaire mother	1	09 January 2006
Phase 6: Study 300 6 month postnatal assessments mother and baby	1	09 January 2006
Phase 7: Study 1500 8 month questionnaire and routine health visitor developmental check (mother)	1	09 January 2006
Phase 7: Study 1500 8 month questionnaire (father)	1	09 January 2006

Research governance approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final research governance approval from the R&D Department for the relevant NHS care organisation.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

NHS
National Research Ethics Service

Cheshire Research Ethics Committee
Western Cheshire PCT
1829 Building
Countess of Chester Health Park
Liverpool Road
Chester
CH2 1HJ

Tel: 01244 650334
Fax: 01244 650333

20 July 2007

Professor Jonathan Hill
Professor of Child and Developmental Psychiatry
Mulberry House, Alder Hey Hospital
Eaton Road
LIVERPOOL
L12 2AP

Dear Professor Hill

Study title: **The Wirral Child Health and Development Study**
REC reference: **05/Q1506/107**
Amendment number: **1**
Amendment date: **31 May 07**

The above amendment was reviewed at the meeting of the Sub-Committee of the REC held on 18 July 2007.

Ethical opinion

The members of the Committee present gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notice of Substantial Amendment (non-CTIMPs)	1	31 May 2007

Membership of the Committee

The members of the Committee who were present at the meeting are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

05/Q1506/107:	Please quote this number on all correspondence

Yours sincerely

Mr Robert Emmett
Committee Co-ordinator

E-mail: rob.emmett@wcheshirepct.nhs.uk

Enclosures List of names and professions of members who were present at the meeting and those who submitted written comments

PRAMMS (T1-T4 - NIMHANS Only)

Bruhath Bengalore Mahanagara Palike
Banashankari Referral Hospital

NO:- BSKRH/MS/PR/~~41~~ 41(a)/14-15. Date :- 22/05/2014

OFFICE MEMORANDUM

Sub :- Granting permission to conduct a study among antenatal mothers – reg.
Ref : - Your requisition letter dated 12/05/2014.

* * *

With reference to the above mentioned subject, Bruhath Bangalore Mahanagara Palike is pleased to grant permission to carry out a prospective study titled "PREVALENCE OF ANXIETY & DEPRESSION AMONG PREGNANT WOMEN & ASSOCIATION WITH ANTENANTAL HEALTH CARE UTILIZATION, ANTENATAL HEALTH BEHAVIOUR AND WITH PREGNANCY OUTCOMES" by your institution faculty at Banashankari Referral Hospital subject to the following terms and condition.

1) No inconvenience should be caused to the hospital inpatients & outpatients.
2) No fees / renumeration should be collected from the study subjects.
3) No damage to be done to BBMP property and equipments.
4) Data collected from antenatal mothers pertaining to the study, to be shared with BBMP authorities every quarter.

As approved by (Chief Health Officer)

Superintendent
Banashankari Refferral Hospital
Bruhat Bengalore Mahanagara Palike
Medical Superintendent
Banashankari Referral Hospital, B.B.M.P.
Bangalore-560 070

27/5/14

To,
Prof. Prabha. S. Chandra,
Principal Investigator,
Department of Psychiatry,
NIMHANS, Bangalore.

Copy To,
1) Chief Health Officer, BBMP for kind information.
2) Office Copy

NATIONAL INSTITUTE OF MENTAL HEALTH AND NEURO SCIENCES
(DEEMED UNIVERSITY)
P.B. NO. 2900, HOSUR ROAD, BANGALORE - 560 029 (INDIA)

Dr. G.S. UMAMAHESWARA RAO,
Professor of Neuroanaesthesia,
Dean-Clinical Neurosciences & Member-Secretary
NIMHANS ETHICS COMMITTEE

Off : 26995004/26564222
Telex: 0845-2186 NIMH IN
Telegram: NIMHANS
Fax: 91-80- 26564830/26566811
email: gsuma123@yahoo.com

No. NIMHANS/85TH IEC/2013 Date: 22.6.2013

1. **Name of the investigator:** Dr. Prabha S. Chandra, Professor, Dept. of Psychiatry

2. **Reference number of the investigator:** Letter dated 28th February, 2013

3. **Title of the proposal submitted for ethical clearance:** **RESEARCH PROJECT FUNDED BY THE INDIAN COUNCIL OF MEDICAL THE RESEARCH(ICMR), NEW DELHI**
Title of the Research Project:

"Prevalence of Anxiety and Depression during pregnancy and association with antenatal healthcare utilization, health behavior and with pregnancy outcomes – prospective study in an urban slum"

4. **Documents submitted:**
   a. Covering letter
   b. Summary Sheet
   c. Research Project Proposal
   d. Consent form (Declaration and Attestation form) duly signed by investigators/collaborators, Head of the department, Head of the Institute
   e. Consent Form (Informed Consent Form) in English and Kannada languages
   f. Undertaking by the investigator
   g. Other information if any
   h. Brief CV of the PI and Co-investigators

5. **Nature of the proposal submitted for ethical clearance:** Research Project

6. **Date of Ethics Committee meeting held:** 6th April, 2013

7. **NIMHANS IEC No. of the proposal: (Please quote this number in all the future correspondence related to this study)** Item No.VII, Sl.No.7.03, Behavioural Science

8. **Members who were present when the proposal was reviewed:**

   1. Dr. A. Jagannatha Rao, DST Rajaramanna Fellow, Member (Scientist), Officiating Chairman, NIMHANS Ethics Committee, Department of Biochemistry, Indian Institute of Science, Bangalore – 560 012

Contd...

2. Dr. Geetha, Professor of Pharmacology, Member (Pharmacologist), NIMHANS Ethics Committee, Dept. of Pharmacology, Bangalore Medical College, Fort, Bangalore – 560 002
3. Dr. O.V. Nandimath, Assoc. Professor, Member, NIMHANS Ethics Committee, National Law School of India University, Bangalore – 560 072
4. Shri. Kishore S.Rao, Managing Trustee, Karunashraya Bangalore Hospice Trust (Community Representative), Member (Lay Person), NIMHANS Ethics Committee, No.208, 5th Main, 11th Cross, Haritha Apartments, Malleswaram, Bangalore – 560 003.
5. Dr. Paritosh Pandey, Assoc. Prof. of Neurosurgery, Member, NIMHANS Ethics Committee, NIMHANS, Bangalore – 560 029
6. Dr. Arun Kumar Gupta, Professor and Head, Dept. of Neuroimaging and Interventional Radiology, (NIIR), Member, NIMHANS Ethics Committee, NIMHANS, Bangalore – 560 029
7. Dr. Jagadisha.T., Addl. Prof. of Psychiatry, Member, NIMHANS Ethics Committee, NIMHANS, Bangalore – 560 029
8. Dr. Thennarasu, Addl. Prof. of Biostatistics, Member, NIMHANS Ethics Committee, NIMHANS, Bangalore – 560 029
9. Dr. Ahalya Raguram, Prof. of Clinical Psychology, Member, NIMHANS Ethics Committee, NIMHANS, Bangalore – 560 029
10. Dr. Shoba Srinath, Prof. & Head, Dept. of Child and Adolescent Psychiatry, Dean (Behavioural Sciences), Member, NIMHANS Ethics Committee, NIMHANS, Bangalore – 560 029
11. Dr. D.K. Subbakrishna, Prof. & Head, Dept. of Biostatistics, Dean (Basic Sciences), Member, NIMHANS Ethics Committee, NIMHANS, Bangalore – 560 029
12. Dr. G.S. Umamaheswara Rao, Prof. of Neuroanaesthesia, Dean (Clinical Neurosciences) and Member – Secretary, NIMHANS Ethics Committee, NIMHANS, Bangalore – 560 029

9. Clear statement of the decision reached:

At the Ethics Committee Meeting held on 6th April, 2013 Committee reviewed the research project and study related documents and discussed the ethical issues involved. After consideration, Committee decided to approve the research project in-principle.

A letter to this effect was sent to you seeking certain clarifications/documents vide letter dated 6.5.2013. In response to this, you have submitted required clarifications/documents vide letter dated 7th June, 2013. Hence, the Research Project and study related documents are approved with respect to ethical aspects.

Decision: Research Project and study related documents are approved with respect to ethical aspects.

DR. G.S. UMAMAHESWARA RAO
DEAN (CLINICAL NEUROSCIENCES) AND
MEMBER – SECRETARY, NIMHANS ETHICS COMMITTEE

Copy to: A.A.O (Project) – for kind information with a copy of the project proposal

NATIONAL INSTITUTE OF MENTAL HEALTH AND NEURO SCIENCES
(INSTITUTE OF NATIONAL IMPORTANCE)
P.B. NO. 2900, HOSUR ROAD, BENGALURU - 560 029 (INDIA)

Dr. B.N. GANGADHAR
Professor of Psychiatry
Dean-Behavioural Sciences & Member-Secretary
NIMHANS ETHICS COMMITTEE

Off : 26995004
Fax: 91-80- 26564830/26566811
Email: bng@nimhans.kar.nic.in
kalyanybg@yahoo.com

No. NIMHANS/DO/98TH IEC/2015

Date: 2.7.2015

1. **Name of the investigator:** Dr. Prabha S Chandra, Professor, Dept. of Psychiatry

2. **Reference number of the investigator:** Letter dated 29.04.2015

3. **Title of the proposal submitted for ethical clearance:** **RESEARCH PROJECT TO BE FUNDED BY INDIAN COUNCIL OF MEDICAL RESEARCH, GOVT. OF INDIA, NEW DELHI**
Title of the Research Project:
"Early psychosocial predictors of child mental health: longitudinal study of shared and distinctive risk and protective factors in UK & India"

4. **Documents submitted:**
   a. Covering letter
   b. Summary Sheet
   c. Research Project Proposal
   d. Consent form (Declaration and Attestation form) duly signed by investigators/collaborators, Head of the department, Head of the Institute
   e. Authorization/sanctioning letter (finance sanction) from sponsoring agency: Nil
   e. Consent Form (Informed Consent Form) for participants (3 types) in English and Kannada languages.
   f. Consent form for carrying out the required investigations from the concerned department(s), if applicable: No
   g. Undertaking by the investigator
   h. Other information if any: Nil
   i. Brief CV of the PI and Co-investigators

5. **Nature of the proposal submitted for ethical clearance:** Research Project

6. **Date of Ethics Committee meeting held:** 20th June, 2015

7. **NIMHANS IEC No. of the proposal: (Please quote this number in all the future correspondence related to this study)** Item No.VII, Sl.No.7.05, Behavioural Sciences

8. **Members who were present when the proposal was reviewed:**

1. Dr. Anura Vishwanath Kurpad (MBBS, MD), Professor of Physiology, Chairman (Medical Doctor), NIMHANS Ethics Committee, St. John's Medical College, St. John's National Academy of Health Sciences, John Nagar, Sarjapur Road, Bengaluru – 560 034
2. Dr. Udaykumar Ranga, Professor, Member (Scientist), NIMHANS Ethics Committee, HIV-AIDS Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur (PO), Bengaluru – 560 064
3. Dr. Srikala Bharath, Professor of Psychiatry, Member, NIMHANS Ethics Committee, NIMHANS, Bengaluru – 560 029
4. Dr. Pratima Murthy, Professor of Psychiatry, Member, NIMHANS Ethics Committee, NIMHANS, Bengaluru
5. Dr. Rita Christopher, Professor of Neurochemistry, Member, NIMHANS Ethics Committee, NIMHANS, Bengaluru – 560 029
6. Dr. N. Shivashankar, Professor of Speech Pathology and Audiology, Member, NIMHANS Ethics Committee, NIMHANS, Bengaluru – 560 029
7. Dr. V. Ravi, Professor of Neurovirology, Dean (Basic Sciences) Member, NIMHANS Ethics Committee, NIMHANS, Bengaluru – 560 029
8. Dr. M. Jayaram, Professor of Speech Pathology and Audiology, Dean (Neuro Sciences), Member –Secretary In-charge, NIMHANS Ethics Committee, NIMHANS, Bengaluru – 560 029

9. **Clear statement of the decision reached:**
Comments:

Action taken by the Member-Secretary in giving provisional approval to the above titled project proposal was ratified. The Committee noted the urgency for giving provisional approval and ratified the decision. The final approval to the research project is to be given.

Decision: The Research Project and study related documents are approved with respect to ethical aspects.

DR. B.N. GANGADHAR
DEAN (BEHAVIOURAL SCIENCES) AND
MEMBER – SECRETARY, NIMHANS ETHICS COMMITTEE

Copy to: A.A.O (Project) – for kind information with a copy of the project proposal

Email sent 1st March 2016 University of Liverpool approval:

Dear Helen,

I am pleased to inform you that your application for recognition of external ethics committee approval has been approved. Details and conditions of the approval can be found below.

Ethics reference number: RETH001024

Review type: Recognition of external ethics committee approval

Title of study: Early psychosocial predictors of child mental health: longitudinal study of shared and distinctive risk and protective factors in UK & India (Early psychosocial predictors of child mental health)

Principal Investigator: Dr Helen Sharp

Co-Applicants: Professor Atif Rahman
Professor Jonathan Hill
Professor Andrew Pickles
Professor Prabha Chandra
Professor Shoba Srinath
Professor Geetha Desai
Professor John Vijaysagar
School/Institute: Psychological Sciences, Institute of Psychology, Health and Society

First reviewer: Professor Graham Kemp

Approval date: 01/03/2016

The application was APPROVED subject to the following conditions:

Conditions

All serious adverse events must be reported to the Subcommittee within 24 hours of their occurrence, via the Research Integrity and Governance Officer (ethics@liverpool.ac.uk).

This approval applies for the duration of the research. If it is proposed to extend the duration of the study as specified in the application form, the Subcommittee should be notified. If it is proposed to make an amendment to the research, you should notify the Committee by following the Notice of Amendment procedure. If the named PI / Supervisor leaves the employment of the University during the course of this approval, the approval will lapse. Therefore please contact the Research Integrity and Governance Officer at ethics@liverpool.ac.uk in order to notify them of a change in PI / Supervisor.

Kind regards,

Mantalena

Mantalena Sotiriadou
Research Ethics and Integrity Officer
Research Support Office
University of Liverpool
Waterhouse Building (2nd Floor, Block C)
3 Brownlow Street
Liverpool
L69 3GL
Email: M.Sotiriadou@liverpool.ac.uk

UKAV (UoL only)

UNIVERSITY OF LIVERPOOL

Central University Research Ethics Committee C

13 January 2020

Dear Prof Sharp

I am pleased to inform you that your application for research ethics approval has been approved. Application details and conditions of approval can be found below. Appendix A contains a list of documents approved by the Committee.

Application Details

Reference:	5741
Project Title:	Using anchoring vignettes to explore variations in mothers' rating behaviour in the areas of maternal mental health and child mental health and development in the UK
Principal Investigator/Supervisor:	Prof Helen Sharp
Co-Investigator(s):	Mrs Laura Bozicevic
Lead Student Investigator:	-
Department:	Psychological Sciences
Approval Date:	13/01/2020
Approval Expiry Date:	Five years from the approval date listed above

The application was **APPROVED** subject to the following conditions:

Conditions of approval

- All serious adverse events must be reported to the Committee (ethics@liverpool.ac.uk) in accordance with the procedure for reporting adverse events.
- If you wish to extend the duration of the study beyond the research ethics approval expiry date listed above, a new application should be submitted.
- If you wish to make an amendment to the study, please create and submit an amendment form using the research ethics system.
- If the named Principal Investigator or Supervisor changes, or leaves the employment of the University during the course of this approval, the approval will lapse. Therefore it will be necessary to create and submit an amendment form within the research ethics system.
- It is the responsibility of the Principal Investigator/Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Central University Research Ethics Committee C

ethics@liverpool.ac.uk

0151-795-8355

Appendix - Approved Documents

(Relevant only to amendments involving changes to the study documentation)

The final document set reviewed and approved by the committee is listed below:

Document Type	File Name	Date	Version
Advertisement	UKAV Study Advert	30/09/2019	1
Debriefing Material	UKAV Debrief	30/09/2019	1
Research Tools	Final EPDS Anchoring Vignettes	30/09/2019	1
Study Proposal/Protocol	UKAV Anchoring Vignette Protocol	30/09/2019	1
Questionnaire	Clarification & Online Links	18/10/2019	1
Research Tools	UKAV Risk Assessment - MBD	15/11/2019	1
Questionnaire	Study 1 Full Questionnaire v2 - January 2020	06/01/2020	2
Questionnaire	Study 2 Full Questionnaire v2 - January 2020	06/01/2020	2
Participant Information Sheet	UKAV Information Sheet - EPDS v3 - January 2020	06/01/2020	3
Participant Information Sheet	UKAV Information Sheet - CBCL v3 -January 2020	06/01/2020	3
Participant Consent Form	UKAV Consent Form - EPDS v2 - January 2020	06/01/2020	2
Participant Consent Form	UKAV Consent Form - CBCL v2 - January 2020	06/01/2020	2
Research Tools	Final Male UK AVs for CBCL - January 2020	06/01/2020	2
Research Tools	Final Female UK AVs for CBCL - January 2020	06/01/2020	2

Central University Research Ethics Committee C

25 March 2020

Dear Prof Sharp,

I am pleased to inform you that the amendment to your study has been approved. Amendment details and conditions of approval can be found below. If applicable, Appendix A contains a list of documents approved by the Committee.

Amendment details

Reference:	5741 (amendment)
Project Title:	Using anchoring vignettes to explore variations in mothers' rating behaviour in the areas of maternal mental health and child mental health and development in the UK
Principal Investigator:	Prof Helen Sharp
Co-Investigator(s):	Mrs Laura Bozicevic
Student Investigator(s):	-
Department:	Psychological Sciences
Approval Date:	25/03/2020

The amendment was **APPROVED** subject to the following conditions:

Conditions of approval

Please note: this approval is subject to the restrictions laid out in the Policy on research involving human participants in response to COVID-19. Therefore all face-to-face contact with human participants for the purpose of research should be halted until further notice; unless the study qualifies as one of the exceptions specified in the Policy and has been discussed with Research Ethics and Integrity team.

- All serious adverse events must be reported to the Committee (ethics@liv.ac.uk) in accordance with the procedure for reporting adverse events.
- If it is proposed to make further amendments to the study, please create and submit an amendment form within the research ethics system.
- It is the responsibility of the Principal Investigator or Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Central University Research Ethics Committee C

ethics@liverpool.ac.uk

0151-795-8355

Appendix - Approved documents

If applicable, the final document set reviewed and approved by the committee is listed below:

Document Type	File Name	Date	Version
Default	UKAV Information Sheet - EPDS v4 - March 2020	23/03/2020	4
Default	UKAV Information Sheet - CBCL v4 - March 2020	23/03/2020	4

EPDS Validation Study (NIMHANS only – Kannada)

ಬೃಹತ್ ಬೆಂಗಳೂರು ಮಹಾನಗರ ಪಾಲಿಕೆ

ಆರೋಗ್ಯಾಧಿಕಾರಿ(ದಕ್ಷಿಣ)ರವರ ಕಛೇರಿ, ಮಹಾನಗರ ಪಾಲಿಕೆ ವಾಣಿಜ್ಯ ಸಂಕೀರ್ಣ ಕಟ್ಟಡ, 2ನೇಮಹಡಿ, 9ನೇಮುಖ್ಯರಸ್ತೆ, 9ನೇಅಡ್ಡರಸ್ತೆ, 2ನೇಬ್ಲಾಕ್ ಜಯನಗರ,ಬೆಂ-11.

ಸಂಖ್ಯೆ:ಆ.ಅ(ದ)/ಪಿಆರ್/ಸಿಎಚ್ /19-20, ದಿನಾಂಕ:-08-08-2019.

ಗೆ,

Dean Cum Director,
National Institute of Mental Health,
& Neuro Sciences
Bangalore-560002.

ವಿಷಯ:- Bangalore Child Health and Development Study (Bchads)ಗಾಗಿ 150 Postpartum ಬಗ್ಗೆ ತರಬೇತಿ ಪಡೆಯುವ ಕುರಿತು.

ಉಲ್ಲೇಖ:- 1. ಡಾ॥ಪ್ರಭಾ.ಎಸ್.ಚಂದ್ರ, ರವರ ಮನವಿ ಪತ್ರದ ದಿನಾಂಕ:-19-07-2019.
2. ಜಂ.ಆ(ದ)ಪಿಆರ್/314/2019-20, ದಿನಾಂಕ:-02-08-2019.

--***

ಮೇಲ್ಕಂಡ ವಿಷಯಕ್ಕೆ ಸಂಬಂಧಿಸಿದಂತೆ, ಡಾ॥ಪ್ರಭ.ಎಸ್.ಚಂದ್ರ, Professor of Phychiatry, Nimhans ರವರು ಪತ್ರ ನೀಡಿ Bangalore Child Health and Development Study (Bchads)ಗಾಗಿ 150 Postpartum ಮಹಿಳೆಯರನ್ನು (Upto 6 months post-delivery) 03 ತಿಂಗಳ ಅವಧಿಗೆ ಬಿಟಿಎಂ ವಲಯದ ತಾವರೆಕೆರೆ ಆರೋಗ್ಯ ಕೇಂದ್ರದಲ್ಲಿ ಶಿಕ್ಷಣ ಪಡೆಯಲು ಅನುಮತಿ ಕೋರಿ ಉಲ್ಲೇಖ (1) ರಂತೆ ಮನವಿ ಸಲ್ಲಿಸಿರುತ್ತಾರೆ.

ಹಾಗೂ ಉಲ್ಲೇಖ (2)ರ ಜಂಟಿ ಆಯುಕ್ತರು (ದಕ್ಷಿಣ) ರವರ ಆದೇಶದಂತೆ Bangalore Child Health and Development Study (Bchads)ಗಾಗಿ 150 Postpartum ಮಹಿಳೆಯರನ್ನು (Upto 6 months post-delivery) 03 ತಿಂಗಳ ಅವಧಿಗೆ ಬಿಟಿಎಂ ವಲಯದ ತಾವರೆಕೆರೆ ಆರೋಗ್ಯ ಕೇಂದ್ರದಲ್ಲಿ ತರಬೇತಿ ಪಡೆಯಲು ಅನುಮತಿ ನೀಡಲಾಗಿದೆ.

ಆರೋಗ್ಯಾಧಿಕಾರಿ (ದಕ್ಷಿಣ)
ಬೃಹತ್ ಬೆಂಗಳೂರು ಮಹಾನಗರ ಪಾಲಿಕೆ

ಪ್ರತಿಯನ್ನು:-

1) ಜಂಟಿ ಆಯುಕ್ತರು (ದಕ್ಷಿಣ) ರವರ ಅವಗಾಹನೆಗಾಗಿ.
2) ಮುಖ್ಯ ಆರೋಗ್ಯಾಧಿಕಾರಿಗಳು (ಸಾರ್ವಜನಿಕ ಆರೋಗ್ಯ) ರವರ ಅವಗಾಹನೆಗಾಗಿ.
3) ಆರೋಗ್ಯ ವೈದ್ಯಾಧಿಕಾರಿ (ಬಿಟಿಎಂ ಲೇಔಟ್) ವಲಯ ರವರ ಮಾಹಿತಿಗಾಗಿ.
4) ವೈದ್ಯಾಧಿಕಾರಿ, ತಾವರೆಕೆರೆ ಆರೋಗ್ಯ ಕೇಂದ್ರ ರವರ ಮುಂದಿನ ಸೂಕ್ತ ಕ್ರಮಕ್ಕಾಗಿ.
5) ಕಛೇರಿ ಕಡತಕ್ಕೆ.

Appendix 2: Participant information sheets

WCHADS – Extensive Phases 1, 3, 5, & 7.

Version 3. March 2007 Mother Information Sheet, Study 1500 – Phases 1,3,5 &7

MANCHESTER 1824

The University of Manchester

Wirral University Teaching Hospital NHS
NHS Foundation Trust

UNIVERSITY OF LIVERPOOL

Study Base:
The Lauries Centre, 142 Claughton Road,
Birkenhead, Wirral, CH41 6EY
Freephone: 0800 051 7597
(from a mobile) 800 051 7597
Text: 07956 297412

Parent Information Sheet (Mother)– Study 1500

Title of study : The Wirral Child Health and Development Study

Investigators: Jonathan Hill, Helen Sharp, Andrew Pickles, Gill Lancaster
Research Staff: Karen Lunt, Carol Bedwell, Belinda Thompson, Julie Carlisle, Kate Marks, Kate Marshall,Liz Green, Florin Tibu, Jo Roberts, Jenny Lee, Nichaela Broyden, Carol Sadler, Jeanette Appleton

You are being invited to take part in a research study. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part. Thank you for reading this.

What is the study about?
We would like to invite you to participate in a new study of children's early development from birth to their first birthdays. This study is based at the Universities of Liverpool and Manchester. It is part of a programme of research into how children learn how to behave with other people, and why some children have difficulties controlling their behaviours. In order to fully understand this we need to measure the early development of children in many different ways. The aim of the study is to find out about the effects of many different forms of stress on parents and babies during the antenatal period and in the first months after birth. We know that for some parents and children the effects are quite long lasting, and others find ways of coping. We want to understand these processes better so that services to support families experiencing stress can be improved.

Who is being invited to take part?
We are approaching all first time mothers and their partners who are booked into the antenatal clinic at Arrowe Park Hospital over a two year period. It is important that we have participants in the study with low, medium and high levels of stress. If you have agreed to take this letter home a research midwife will contact you at your 20 week appointment or slightly after, to tell you more about the study, answer any questions you have and to invite you to take part.

Do I have to take part?
It will be up to you to decide whether or not you would like to take part. If you agree, and change your mind later, you can withdraw from the study. This will not affect the care you receive.

How often will I be contacted?
We will contact you again six weeks after the birth of your baby, and when your baby is 8 months old. We would also like to contact some mothers more often up to the first birthdays of their children, so that we can ask them more about their lives, and understand better their ways of coping, and assess their babies' health and development in more detail. If you decide to take part, the computer will tell us who to invite for the additional contacts after we have entered the information you provide now. If your name does come up we hope very much that you will be able to help us, but at this stage we are only asking you to participate now and at 6 weeks and eight months.

What will I be asked to do at each time point?
During your pregnancy we will interview you and ask you to complete some questionnaires about your current health and relationships, and about your expectations of the baby and being a mother. This can be done here at the antenatal clinic or at another clinic on the Wirral or at the study base in the Lauries Centre. It should take about 25 minutes.

We will also ask you for consent for us to have access to your medical records for the pregnancy, the birth, and your new born infant following the birth.

When your baby is 6 weeks old we will send you some short questionnaires about your health, your relationships, and about your baby by post, and ask you to 'Freepost' them back to us.

When your baby is 8 months old we will send you more questionnaires about your health and about your baby, and ask you to return them 'Freepost' to us or return them to your health visitor when you attend for your baby's routine 8 month developmental check-up. We will also ask your health visitor for the results of their 9-12 month assessment of your baby's development.

If you give written consent to take part in this study and you are selected by the computer to be invited for additional contacts, one of the research team named on the front of this information sheet will contact you at home, using the contact details you give to the research midwife. They will only contact you if you agree to it.

How will this information be used?
All information that we receive from you will be treated as strictly confidential, under the guidelines of the Universities of Liverpool and Manchester, the UK Medical Research Council, and the Data Protection Act. Information that we enter on the computer will be identified only by a number. We will report general findings about parents and children, but you or your child will never be identified. The only reason we might have to share information from the study with other people is if there are concerns about you or a child being at risk of serious harm. If that happens we will talk with you first to decide on the best way forward. Concerns like this would be addressed by seeking appropriate forms of help for you and/or following Trust Child Protection Guidelines.

Who is organising and funding the research study?
The study is being run by Professor Jonathan Hill of the University of Manchester and Dr Helen Sharp of the University of Liverpool. The research is funded by the Medical Research Council.

Are there any benefits in taking part in this study?
There are no benefits to you or your child's health in taking part in this study. However we hope that you will feel you are contributing to medical research in a way that will help children and families in the future.

What if something goes wrong?
If you feel you or your child have been harmed by taking part in this research and that the researchers have been negligent or at fault, then you may be able to make a legal claim for compensation to their employer. You might have to pay the legal costs of doing this. However, if you are harmed and the researchers are not at fault, there is no facility for you to make a claim. If you wish to complain or have any concerns about any aspect of the way you have been approached or treated during the course of this study, normal University or National Health Service complaints procedures should be available to you.

Are there any risks to myself or my child taking part in this study?
No, there are no known or likely risks.

Who has reviewed and approved the study?
A team of international experts on child development has reviewed this study for the Medical Research Council. The study has been reviewed and approved by the Research & Development committees of Wirral Hospitals NHS Trust, Wirral PCT and the Cheshire Local Research Ethics Committee.

Can I ask further questions?
When the research midwife meets you, at or after your 20 week scan appointment, she will be very happy to answer any questions you might have. In the meantime, if you would like any more information, please do not hesitate to contact Professor Jonathan Hill, Dr Helen Sharp, or Liz Green on the freephone number shown on the front page.

MANCHESTER 1824

The University of Manchester

Wirral University Teaching Hospital NHS
NHS Foundation Trust

UNIVERSITY OF LIVERPOOL

Study Base:
The Lauries Centre, 142 Claughton Road,
Birkenhead, Wirral, CH41 6EY
Freephone: 0800 051 7597
(from a mobile) 800 051 7597
Text: 07956 297412

Parent Information Sheet – Study 300

Title of study : The Wirral Child Health and Development Study

Investigators: Jonathan Hill, Helen Sharp, Andrew Pickles, Gill Lancaster
Research Staff: Kate Marks, Florin Tibu, Kate Marshall, Melissa Bensinyor, Helen Jones, Liz Green, Nicola Sandman, Alice Hulbert, Kirsty Entwistle, Gemma Culverwell, Louise Fisher, Stuart Kehl, Fay Huntley

When you were pregnant, and again just after your baby was born you kindly helped us with a study that we are conducting designed to understand better how stress affects mothers to be, their partners and their babies, and how good experiences and support can make a difference. We are following 1500 women up to the first birthday of their babies mainly using questionnaires. In addition we are asking 300 to take part in interviews and to agree to us filming their babies during the first year of their life. You are one of the 300 that we would like to see again now that your baby is one year old. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part. Thank you for reading this.

What is the study about?
The aim of the study is to find out about the effects of stress on parents and children during the antenatal period and in the first months after birth. We plan to measure each baby's development and how they interact with their mother in some detail. We believe that for some parents and children the effects are quite long lasting, and others find ways of coping. We want to understand these processes better so that services to support families experiencing stress can be improved.

Our research team is very interested to know more about the genes that influence children's emotions and behaviours. Every child is a unique individual, and that is partly due to the genes that have been passed on from each parent. Genes are like maps inside our bodies that hold information. For example, it is well known that the colour of our eyes depends on our genes. More recently we have learnt much more about how health and behaviour are influenced by genes. This study provides an important opportunity to learn more about the ways in which genes affect the way infants behave and their ability to cope with new situations.

Who is being invited to take part?
The computer chooses the names of women who we approach based on the information they have given about how much stress they may be facing. Because we particularly want to understand about stress in pregnancy the computer is picking more women who are experiencing stress. Your name has been chosen either because you have indicated that you may be dealing with quite a lot of stress or because you have said you are not facing a lot.

Do I have to take part?
It will be up to you to decide whether or not you would like to take part. If you agree, and change your mind later, you can withdraw from the study. This will not affect the care you receive.

How often will I be contacted?
Now that your baby is one year old we would like you and the baby to come to our study centre for about half a day. We are planning further contacts for the future and we hope we will be able to obtain funding to see you again when your baby is around two to two and a half years old.

What will we have to do?

- We would like to see you and your child at the Lauries Centre for half a day.
- We will talk with you about your feelings and experiences since the last visit and audio record our conversation.
- We will ask you about your child's behaviours and emotions. For example we will ask what makes him/her anxious, or angry, or happy, and what he/she likes to do with you. We will audio record this conversation also.
- We would like to make a short video (about 15 minutes) of your baby playing with you with some toys .
- We would like to make a video of how your baby responds to everyday events such as playing with various toys, seeing an unusual character or not being allowed to play with a toy for a short time.
- We would also like to make a video of how your child responds to being separated from you. Some children find this quite hard and others are not worried by it. You will be able to see your child's response and if he or she is distressed by it you will be able to comfort him/her straight away. This experience is designed to mimic or copy natural times at home when you have to separate for a short time, for example while you go briefly into another room.
- We will put two patches on your baby's chest (just as we did when your baby was younger) to record your baby's heart during video recordings of your baby and of the separation and when he/she is with you again.
- We are also going to see whether some babies are more likely to produce the kinds of hormones that help them to deal with challenging situations. To do this, all we have to do is ask your baby to chew on a soft, cotton dental roll, which is completely safe, and will not produce any allergic reactions. This allows us to collect a sample of your baby's saliva, which can then be analysed to measure the hormones. We would like to do this four times, once before, and once after the separation from you, and once before and once after a toy play task.
- We would also like to collect saliva from your baby for DNA analysis using a similar cotton swab.
- We would like to find out about your child's development by giving him/her some puzzles to solve.
- We will weigh your child and measure their height and head size.

Will my expenses be paid?
We will be pleased to organise transport to the interview, or to pay for your transport. We are able to pay up to £30 to compensate you for time lost from home or work or any other expenses incurred from taking part in the study.

How will this information be used?

- We would like to make a video recording of your baby and you so that we go over what has happened in detail afterwards. The recording will be identified only by a number, so that information on it cannot be traced to you. The recording will be kept secure at the university base for up to ten years.
- All information that we receive from you will be treated as strictly confidential, under the guidelines of the Universities of Liverpool and Manchester, the UK Medical Research Council, and the Data Protection Act.
- Information on audio and video recordings, and on paper records, and that we enter on to the computer will be identified only by a number. A list of names and addresses of participants and their case numbers will be kept separately and securely in the university base.
- The genetic samples will be analysed anonymously. No records will be generated that directly link your name, your partner's name, or your child's name to the genetic samples. They will only be analysed for the purpose of this study, and will never be analysed for any other purpose. We will analyse the samples for genes that affect infants' emotions and behaviour, and not for any other purpose. They will not be kept as part of your medical record. All samples will be destroyed after 20 years. The anonymous samples will be analysed by a laboratory technician who is not affiliated with the study, and will have no access to your name, your partner's name, or your child's name.
- We will report general findings about parents and children, and you or your child will never be identified. Reports will only be based on the ratings that we make from the interview and none of what you say will be reported.
- The only reason we might have to share information from the study with other people is if there are concerns about you or a child being at risk of serious harm. If that happens we will talk with you first to decide on the best way forward. Concerns like this would be addressed by seeking appropriate forms of help for you and following Trust Child Protection Guidelines.

Who is organising and funding the research study?
The study is being run by Professor Jonathan Hill of the University of Manchester and Dr Helen Sharp of the University of Liverpool. The research is funded by the Medical Research Council.

Are there any benefits in taking part in this study?
There are no benefits to your or your child's health in taking part in this study. However we hope that you will feel you are contributing to medical research in a way that will help children and families in the future.

What if something goes wrong?
If you feel you or your child have been harmed by taking part in this research and that the researchers have been negligent or at fault, then you may be able to make a legal claim for compensation to their employer. You might have to pay the legal costs of doing this. However, if you are harmed and the researchers are not at fault, there is no facility for you to make a claim. If you wish to complain or have any concerns about any aspect of the way you have been approached or treated during the course of this study, normal University or National Health Service complaints procedures should be available to you.

Are there any risks to myself or my child taking part in this study?
No, there are no known or likely risks.

Who has reviewed and approved the study?
A team of international experts on child development has reviewed this study for the Medical Research Council. The study has been reviewed and approved by the Research & Development committees of Wirral University Teaching Hospital NHS Trust, Wirral Primary Care Trust, Western Cheshire PCT and the Cheshire Local Research Ethics Committee.

Can I ask further questions?
When the researcher meets you they will be very happy to answer any questions you might have. In the meantime, if you would like any more information, please do not hesitate to contact Professor Jonathan Hill, Dr Helen Sharp or Liz Green on the freephone number shown on the front page.

BCHADS

INFORMED CONSENT FORM

Title of the study: Early Psychosocial predictors of child mental health: longitudinal study of shared and distinctive risk and protective factors in UK and India

Information to the participants:

In this study, the researchers are attempting to follow-up children and families through infancy to childhood, to study developmental risk and protective factors for childhood mental health problems and to assess behavioural outcomes.

To gather all this information, the researchers will interview you for a total of five times. The assessments will be as follows

Assessments	**Infant age**
1st Assessment	6 months
2nd Assessment	1year
3rd Assessment	2 years
4th Assessment	2 ½ years
Last Assessment	3 years

The interviews will be carried out over 1 or 2 sessions by the researcher with each session lasting for about half an hour. The interviews will focus on your health and emotional status during the postpartum period, breastfeeding and on your interaction and involvement with the infant. Further, details about the nature of childcare that your baby receives from multiple persons at home will also be discussed in these interviews. The researcher will seek a 5 minute speech sample during the 1st assessment, where you will be asked to speak about your baby and about your engagement with your baby and the same will be audio-recorded. In addition to these, infant health and temperament, behavioural symptoms and cognitive development would also be examined during each of these assessments.

Salivary sample from the infant will be collected at 2 years of age for a DNA assessment.

The information given by you will be kept completely confidential. These sessions will be conducted at the clinic, at the nearest anganwadi centre or at home depending on your convenience.

Information regarding the infant's birth details will be taken from the birth records. All information will be completely confidential and your data will not have any identification except for code numbers i.e your name and address details will not be available to anyone.

Risks from participating in the study
There is no major risk to participating in the study. The only possibility is that you might feel uncomfortable while discussing your emotions and feelings.

Benefits of participating in the study
This study may not benefit you directly but your participation in the study will help us to evolve appropriate intervention strategies in the future for women who may have postpartum anxiety or depression or may experience difficulties in mothering.If we find that you have an emotional issue that requires medication or counselling, we will refer you to an appropriate health professional with your consent. We will also be working closely with your obstetrician and your baby's paediatrician and if we find any health concerns, we will refer you to the service with your consent.

You have the right to withdraw from the study at any point of time. Refusal to participate in the study will not affect your treatment in any way. The information given by you will be used only for the study.

Undertaking by the researcher:
Your consent to participate in the above study is solicited. Your choice to participate in this is entirely voluntary. The data will not identify you in any way except to the researcher. You will have the right to refuse consent or withdraw during any phase of the study without giving any reason. There are no tangible benefits for participation in the study but appropriate interventions will be provided where required. However your participation in this study will be of immense help in understanding ways of helping other women with the similar difficulties. If you have doubts regarding the study, please feel free to clarify the same. You are free to contact the investigator for the further guidance or help if you desire.

Dr. Prabha S. Chandra, Professor, Dept of Psychiatry. Contact No: 08026995272

Consent:
I have been informed about the procedures of the study. I have understood that I have the right to refuse my consent or to withdraw from the study at any point of time. I, ……..the undersigned give my consent to be participant for the study.

Signature of the respondent (Name and address)

Signature of the witness (Name and address)

Signature of the investigator (Name and Designation)

Date

Place:

UKAV

Information Sheet

Title of study: Using anchoring vignettes to explore variations in mothers' rating behaviour in the area of child mental health and development in the UK.

Version Number & Date: Version 4: March 2020

Investigators: Professor Helen Sharp, Professor Jonathan Hill, Professor Andrew Pickles, Professor Prabha Chandra, Dr K. Thennarasu, Dr M. Thomas Kishore, Matthew Bluett-Duncan, Laura Bozicevic.

You are being invited to participate in a research study. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask us (via the email address at the end of this document) if you would like more information or if there is anything that you do not understand. Please also feel free to discuss this with your friends, relatives and GP if you wish. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to. Thank you for reading this.

What is the purpose of the study?

We are interested in how mothers' rating behaviour of the child's emotions and behaviours vary between India and the UK. Rating behaviour refers to how people use response sets when completing questionnaires, for example *"Often True, Sometimes True, Not True."*. Research suggests that the way people use the response options on questionnaires may differ as a result of the social and cultural norms and practices they are exposed to. In turn this can lead researchers to make incorrect conclusions about whether or not there are real differences between countries.

We will compare the responses given in this project to responses given in a separate project in India and calculate a "correction score" that will be applied to data collected by two large child health and development research studies that have taken place in India and the UK. This correction score will remove the influence of any systematic differences in response style. This will help us to understand what proportion of the difference we observe between India and the UK is due to actual differences in child development, and how much is simply the result of cultural differences in how people answer questions. This means we will be able to make accurate comparisons between the two countries and better identify the situations and experiences that may increase or decrease the risk of poor child development, and whether these are the same or different in Western and Southeast-Asian contexts.

Why have I been invited to take part?

We would like to ask you to take part if you are a mother who lives in the UK, you can speak and read English, and you have a child who is 2-3 years old. Each participant should only complete the questionnaire once. You will not be reimbursed for multiple entries.

Do I have to take part?

Participation in this study is completely voluntary. If you do decide to take part, you can withdraw at any point during or following completion of the questionnaire. You do not have to give any

explanation as to why you are withdrawing and this will not affect your rights in any way. Your data will only be able to be identifiable by your Prolific ID number and we will use this to locate and remove any responses that are withdrawn. If you stop the questionnaire part way through, we will treat this as an indication you have withdrawn your consent and none of your responses will be used in the study.

What will happen to me if I take part?

This is an online survey which can be completed on any PC, laptop or mobile device in the place of your choosing. You do not need to come to the university to take part. If you choose to proceed we will ask you to provide some background information about yourself, your child's age, and to fill out a questionnaire about your child's behaviours and emotions. Following this we will show you a number of short descriptions, or 'vignettes', of hypothetical characters and will ask you to rate their symptoms or behaviours. Previous use of the questionnaire and vignettes tells us that the whole survey should take 25 minutes to complete. The same short descriptions, or "vignettes", have already been shown to a group of mothers in India after being translated into their local language.

How will my data be used?

The University processes personal data as part of its research and teaching activities in accordance with the lawful basis of 'public task', and in accordance with the University's purpose of "advancing education, learning and research for the public benefit".

Under UK data protection legislation, the University acts as the Data Controller for personal data collected as part of the University's research. Professor Helen Sharp acts as the Data Processor for this study, and any queries relating to the handling of your personal data can be sent to hmsharp@liverpol.ac.uk.

Further information on how your data will be used can be found in the table below.

How will my data be collected?	Data will be collected using the Qualtrics online survey platform.
How will my data be stored?	Data will be stored on the Qualtrics Secure Server for the duration of the study. When the study is complete it will be transferred to a secure, password protected network drive at the University of Liverpool and deleted from Qualtrics.
How long will my data be stored for?	30 years
What measures are in place to protect the security and confidentiality of my data?	All information that we receive from you will be anonymised and treated as strictly confidential, under the guidelines of the University of Liverpool and the Data Protection Act. This means that your information will only used by members of the research team and scientific research collaborators from other academic institutions approved by us.
Will my data be anonymised?	Yes, all data will be anonymised and identifiable only by your anonymous Prolific ID.

How will my data be used?	Your data will be used as part of group analyses in conjunction with data from 2 other research projects.
Who will have access to my data?	Members of the study research team. Other researchers will have access to the data in a fully anonymised format once it has been uploaded to DataCat (see next section).
Will my data be archived for use in other research projects in the future?	Yes, data will be anonymised and made available for re-use through DataCat (UoL's research data catalogue).
How will my data be destroyed?	Data will be held anonymously on DataCat and registered for deletion after 30 years has elapsed.

Expenses & Payments

We are able to reimburse you with £2.50 via Prolific to compensate you for the time taken to complete the questionnaire.

Will my taking part in the study be kept confidential?

Yes. All data will be stored in an anonymised format and identifiable only by your anonymous Prolific ID.

What will happen to the results of the research study?

Results will be written up as part of a PhD thesis, published in scientific journal articles, and presented at academic conferences.

What are the possible benefits to taking part?

There are no benefits to your health in taking part in this study. However, we hope that you will feel you are contributing to medical research in a way that will help children and families in the future.

What are the possible disadvantages and risks to taking part?

There are no known or likely risks to taking part in this study. The questionnaire is a widely used measure of child behavioural and emotional problems that has been validated within the research community. However, it is possible that some of the items may cause mild distress by raising your awareness of potential difficulties that your own child may have. Items include questions about child aggression, anxiety, depression and attention problems. The vignettes describe fictional characters with some of these symptoms. While these vignettes have been used previously without any known issues, they do have the potential to cause distress if your child is experiencing similar challenges. If this happens you are free to skip a particular item or stop the questionnaire at any point. We will also provide information regarding relevant sources of support at the end of the study.

What will happen if I want to stop taking part?

You can withdraw at any point without explaining why, up until the end of the questionnaire. If you stop the questionnaire part way through, we will treat this as an indication you have withdrawn your consent and none of your responses will be recorded. After you submit your responses, you will not be able to withdraw from the study.

Who has reviewed and approved this study?

The study has been reviewed and approved by the University of Liverpool Ethics Committee.

What if there is a problem?

If there are any problems or if you would like more information before or after taking part please contact either the research student or supervisor, or the University of Liverpool Ethics Team, using the details below:

Primary Investigator: Professor Helen Sharp (hmsharp@liv.ac.uk)

Student Investigator: Matthew Bluett-Duncan (hlmbluet@liv.ac.uk)

Independent Contact: University of Liverpool Ethics Team (email: ethics@liverpool.ac.uk, Phone: 0151 794 8290)

Appendix 3: Systematic Review

Antenatal Main Effects

Author (Year)	Design	Cognitive Development			Language Development		
		Mean Scores (SD)		Main Effect	Mean Scores (SD)		Main Effect
		Depressed	Non-Depressed		Depressed	Non-Depressed	
Bandoli et al., (2016)	RCT	**6m:** 90.9 (6.8) **12m:** 90.8 (10.0)	**6m:** 90.5 (9.2) **12m:** 93.0 (11.6)	**6m**: B =-1.96 (4.44-0.52) NS **12m**: B = -0.16 (3.54-3.22) NS	N/A	N/A	N/A
Breen et al., (2018)	Prospective	**24m:** 83.7 (7.79)	**24m:** 90.0 (6.16)	**24m:** Effect not given, p <0.001.	**24m:** 83.78 (14.09)	**24m:** 87.12 (9.89)	**24m:** NS
Donald et al., (2019)	Prospective	Not Given	Not Given	**24m:** β = −1.03; (−1.04, −0.12), p = 0.027.	Not given	Not given	Not given
Lin et al., (2017)	Prospective	N/A	N/A	N/A	Not given	Not given	**24-30m:** β = -13.18 (-24.14, -2.22), p = 0.012.
Munoz-Rocha et al., (2018)	Prospective	Not given	Not given	**Model 1** **24-30m:** β = -2.40 (0.8), p < 0.01 **Model 2:** **24-30m:** β = -2.2 (1.1), p = 0.06.	Not given	Not given	**Model 1** **24-30m:** β = -2.47 (0.9), p = 0.01 **Model 2** **24-30m:** β = -2.17 (1.2), p = 0.08.

Murray et al., (2016)	RCT	81.4 (CI = 78.6, 84.3)	84.8 (CI = 83.4, 86.2)	**Model 1** **18m** F = 4.4, p = 0.04 **Model 2** **18m** F = 3.1, p =0.08	N/A	N/A	N/A
Rotheram-Fuller et al., (2018)	RCT	**OS – 36m:** AND: 1.27 (SE=0.24) PND: 0.98 (SE=0.18) A/PND: 1.92 (0.21)	**OS – 36m:** Never: 1.44 (0.15)	**OS – 36m:** All groups NS	**N/A**	N/A	N/A
		SS – 36m: AND: 6.11 (0.68) PND: 5.47 (0.52) A/PND: 5.32 (0.62)	**SS – 36m:** Never: 5.91 (0.45)	**SS – 36m:** All groups NS.			
		STS – 36m: AND: 5.21 (0.64) PND: 4.29 (0.49) A/PND: 4.17 (0.60)	**STS – 36m:** Never: 4.30 (0.42)	**STS – 36m:** All groups NS			
		PPVT – 36m: AND: 19.66 (0.79) PND: 19.20 (0.59) A/PND: 19.89 (0.73)	**PPVT – 36m:** Never: 19.79 (0.53)	**PPVT – 36m:** All groups NS.			
Tran et al., (2013)	Prospective	**6m:** 97.92 (14.13)	**6m:** 100.03 (12.82)	**6m:** B = -4.80, (95% CI: -9.40, -0.20), p < .05.	N/A	N/A	N/A

Key: OS = Operation Span; SS = Silly Sounds; STS = Something's the Same; PND = Postnatal Depression; PPVT = Peabody Picture Vocabulary Test; AND = Antenatal Depression; A/PND = Antenatal and Postnatal Depression.'

Postnatal Main Effects

Author (Year)	Design	Cognitive Development			Language Development		
		Mean Scores (SD)		Main Effect	Mean Scores (SD)		Main Effect
		Depressed	Non-Depressed		Depressed	Non-Depressed	
Ali et al., (2013)	Quasi-Experimental	Not given	Not given	**6m:** OR= 3.3 (95% CI: 1.1, 9.9)* **12m:** OR= 6.8 (95% CI: 3.0, 15.7)*	Not given	Not given	NS (effect not given)
Black et al., (2007)	RCT	Not given	Not given	**12m:** B = .09, *p* > 0.05.	**N/A**	N/A	**N/A**
Familiar et al., (2018)	Prospective	Not given	Not given	**12m:** B = -2.49, (-5.86, 0.88), p = 0.15.	N/A	N/A	N/A
Galler et al., (2000)	Prospective	Not given	Not given	**3m:** F(3, 78)=2.09; p<0.02. **6m:** NS (effect not given)	N/A	N/A	N/A
Garman et al., (2019)	RCT-Control Arm	Early: 10.24 (2.72) Late: 9.59 (3.30) Chronic: 9.00 (1.84)	Chronic Low: 10.14 (3.03)	**BSID - 18m:** β = 0.08, p=0.91** **OS - 36m:** β =-0.29, p = 0.47** **SS – 36m:** β = -0.75, p = 0.54** **STS – 36m:** β =-0.11, p = 0.89**	N/A	N/A	N/A
Hamadani et al., (2012)	Prospective	99.7 (10.8)	100.6 (12.1)	**12m:** NS (no effect given)	N/A	N/A	N/A

Patel et al., (2003)	Prospective	86.4 (84.1, 88.8)	90.3 (87.7, 92.9)	**6m:** OR = 3.3 (1.2, 8.8), p = 0.02.*	N/A	N/A	N/A
Quevedo et al., (2012)	Prospective	N/A	N/A	N/A	**Postpartum**: 107.24 (16.48) **Current**: 105.95 (14.08) **Postpartum & Current**: 97.43 (15.40)	None: 108.59 (17.00)	**12m:** B = -2.87 (-5.01; -0.64) p = 0.01***
Tran et al., (2013)	Prospective	102.88 (14.76)	98.54 (13.07)	**8w:** B = 1.26 (-1.02; 3.54), *p* > 0.05. **6m:** B = 1.76, (-.09; 3.62), *p* > 0.05.	N/A	N/A	N/A

* ORs given for likelihood of developmental delay.

** Effect given for early postpartum depressed group only.

*** Effect given for no depression vs any depression. Post-hoc tests revealed effect of postpartum & current group only.

Key: BSID = Bayley Scales of Infant Development; OS = Operation Span; SS = Silly Sounds; STS = Something's the Same.

Search Strategy for each database		Participants	Exposure	Outcome Population	Outcome	Context
Key concepts		**Mothers**	**Depression**	**Infants**	**Cognition**	**Low and middle-income countries**
Free text terms / natural language terms **Search terms identical for:** • **PubMed** • **PsycInfo** • **CINAHL**		Mothers Maternal Perinatal Peripartum Prenatal Antenatal Antepartum Pregnancy Pregnant Trimester Postnatal Postpartum Puerperal Puerperium Post-birth	Depression Depressed Depressive symptoms Depression symptoms Depressive disorder Affective disorder	Infant Infants Child Children Infancy Baby Toddler Newborn Childhood Pre-school	Cognition Cognitive Language IQ Intelligence Memory Perception Learning Problem solving Metacognition Social cognition DQ Communication Executive function Attention Concentration	Low income population Low Income Country/ies Middle income population Middle Income Country/ies Low and middle income population Low and middle income country/ies Developing Country/ies Developing nations Third world Poverty LMIC LAMIC Africa Asia South America Central America South Asia Middle East *+ list of all LAMICs*
Controlled vocabulary	**PubMed**	Peripartum Postpartum period (noexp) Prenatal care	Postpartum depression Depression Postpartum depression Depressive disorder	Infant Infant, Newborn (noexp) Child, preschool	Cognition Intelligence Learning	Asia Africa South America Central America

terms / Subject terms		Postnatal care		Child development	Poverty (encompasses LAMIC)	
	PsycInfo	Postnatal period Antepartum period Perinatal period Mothers Postpartum depression	Depression Postpartum depression Affective disorders	Infant development Early childhood development	Cognition Delayed development Intelligence Infant development Cognitive development Cognitive ability	Poverty Poverty areas Developing countries Emerging economies
	CINAHL	Postnatal period Postnatal care Prenatal care Perinatal care Mothers Expectant mothers	Depression Postpartum depression Affective disorders	Infant Infant, newborn Child, preschool	Infant development Child development Language development Cognition Mental processes (Exp)	Low and middle income countries Developing countries Africa Asia Indian Ocean Islands Pacific Islands Central America South America Latin America West Indies Poverty Poverty Areas

Appendix 4: EPDS Validation Study + Measures

EPDS Validation Study

Summary

The objective of this study was to assess the validity of a Kannada version of the Edinburgh Postnatal Depression Scale (EPDS) and to establish separate cut-off points for probable depression in the antenatal and postnatal periods. Separate samples of 150 pregnant women and 98 mothers of young infants completed the EPDS and a short demographic interview. The presence or absence of a clinical diagnosis of a major depressive episode was confirmed by the Mini International Neuropsychiatric Interview (MINI), administered by a psychiatrist or psychiatric social worker blind to the EPDS response. The ROC curve analysis revealed that the EPDS was a fair quality instrument during the antenatal period (AUC = .734) and a score of 3 or more was selected as the optimal cut-off to screen for probable antenatal depression in women (sensitivity = 62.5%, specificity = 73.9%). Due to patterns of reporting in the BCHADS sample, this cut-off was also applied to the 8-week postnatal assessment. The ROC curve analysis during the postnatal period indicated that the EPDS is an excellent quality instrument at this time-point (AUC = 0.945) and a score of 10 or more was selected as the optimal cut-off to screen for probable postnatal depression (sensitivity = 95.2%, specificity = 83.1%). This cut-off was applied to the 6, 12 and 24 month assessments in BCHADS.

1. Introduction

A recent meta-analysis of postnatal depression prevalence showed that 29 out of 38 studies conducted in India used the Edinburgh Postnatal Depression Scale (EPDS) (Cox, Holden, & Sagovsky, 1987), either independently or alongside another measure. However, only 8 of these studies reported a validated cut-off point for probably depression while the majority utilised thresholds validated in western populations. Evidence suggests that the EPDS functions differently in different settings, with thresholds especially likely to vary between different cultures (Gibson, McKenzie-McHarg, Shakespeare, Price, & Gray, 2009). Therefore, it is important to validate each adaptation within its context.

The EPDS has previously been translated into Kannada and validated in a rural population (Fernandes et al., 2011). The authors report a cut-off score of 13 or more (sensitivity=100%, specificity=84.90%, AUC=0.95) for detecting prenatal depression but note the findings may not be generalisable to an urban setting where different cut-offs may be expected. In view

of this, the current study recruited two independent samples, one antenatal and one postnatal, to validate a Kannada version of the EPDS in an urban setting.

2. Method

This study used a cross-sectional design and was carried out between September 2015 and December 2015 (antenatal sample) and from August 2019 to October 2019 (postnatal sample).

2.1. Antenatal Sample

A consecutive sample of 150 women were recruited from the antenatal clinic at the Banashankari Urban Primary Health Centre in Bangalore. Pregnant women living in low income areas of urban Bangalore (India) and registered with the Antenatal clinic at a Government Referral Hospital (GRH) in South Bangalore were potentially eligible to participate. Women who had a major mental illness such as psychosis or a bipolar disorder, who were identified to have major health complications during the current pregnancy or were currently using alcohol or other psychoactive substances were excluded.

2.2. Postnatal Sample

A consecutive sample of 98 mothers were recruited from an antenatal/maternity clinic in South Bangalore when attending routine immunisation appointments with their infants. Mothers living in low income areas of urban Bangalore (India) and registered with the Antenatal clinic at a Government Referral Hospital (GRH) in South Bangalore were potentially eligible to participate. Women who had a major mental illness such as psychosis or a bipolar disorder, who were identified to have major health complications during the current pregnancy or were currently using alcohol or other psychoactive substances were excluded.

2.3. Measures

Edinburgh Postnatal Depression Scale (EPDS): The EPDS is a 10-item Likert scale self-report instrument designed to detect postnatal depression by focusing on the cognitive and affective aspects of depression rather than somatic symptoms. Each item addresses a distinct symptom of postnatal depression and is rated using a unique set of response items, scored 0-3. The instrument gives a total score of 0-30, with a higher score indicating greater distress.

The EPDS was translated into the local language (Kannada) following World Health Organisation guidelines for measure translation. Forward translation was done by a researcher from the team who was a native speaker of Kannada with bilingual proficiency in Kannada and English. The translated version was then reviewed by experts in the field to ensure the appropriateness of the terms and phrases used in the questions and revisions were made based on their suggestion. This version was then back-translated to English by a translator with bilingual proficiency. The original English questionnaire, the translation in Kannada and the back-translated version were compared and reviewed to identify any differences and appropriate modifications were made. Finally, a pilot study was conducted with 10 pregnant mothers from the target population. Final modifications were made to the phrasing of items 1, 2, 4, 5 and 6 based on pilot feedback.

Mini International Neuropsychiatric Interview (MINI): The reference standard diagnostic interview was the MINI version 5.0.0 (Lecrubier et al., 1997), a short structured diagnostic interview for DSM-IV and ICD-10 psychiatric disorders. The MINI has been validated against the SCID for DSM-III R and the CIDI for ICD-10 (Sheehan, et al. 1997; Spies et al. 2009), and has been used extensively in field research in India (Fernandes et al., 2011).

2.4. Procedure

In both studies the women first completed the EPDS and a short demographic interview. Although the EPDS was originally designed to be self-administered, low literacy levels in the sample meant that the scale had to be researcher administered, with researchers reading out each item and then the set of responses to be selected by participants. The MINI interview was then conducted in a separate room by either a psychiatrist or psychiatric social worker who was blind to the responses given on the EPDS.

2.5. Analysis

All analyses were carried out using Statistical Package for Social Sciences (SPSS) version 24 for Windows. Descriptive statistics regarding the sociodemographic characteristics and mean EPDS scores were produced first and are presented alongside corresponding statistics from the main BCHADS sample. The sample was categorised into cases and non-cases based on a MINI diagnosis of major depressive disorder (MDD). Receiver-operating characteristic curves were then calculated to determine sensitivity, specificity, and the area under the curve (AUCs).

3. Results

3.1. Sample Characteristics

Sample demographics for the antenatal and postnatal validation cohorts are presented alongside the baseline demographics for the BCHADS cohort in table 1.

Table 1: Comparison of sample demographics

Characteristic		Antenatal Validation	Postnatal Validation	BCHADS Baseline
Maternal Age		22.77 (3.21)	26.52 (4.60)	22.92 (3.66)
Maternal Education (years)		9.89 (2.76)	11.16 (3.53)	9.81 (2.60)
Occupation (Employed)		7.3%	9.2%	13.8%
SES	BPL/LSES	8.8%	5.1%	11.3%
	Upper-LSES	52.7%	26.5%	44.8%
	MSES	38.5%	51%	41.2%
	USES	0.0%	14.3%	2.8%
Religion	Hindu	78.0%	67.3%	83.5%
	Muslim	21.3%	26.5%	15.4%
	Other	0.7%	6.1%	1.1%

3.2. Depression Prevalence

The mean EPDS score during the antenatal period was 2.93 (SD = 5.32). 16 women met diagnostic criteria for a major depressive episode on the MINI, giving an antenatal prevalence of 10.6%.

The mean EPDS score during the postnatal period was 7.61 (SD = 5.37). 21 women met diagnostic criteria for a major depressive episode on the MINI, giving a postnatal prevalence of 21.4%.

The current mean EPDS scores are presented alongside mean EPDS scores from each BCHADS study phase in table 2. Due to similarities in mean scores between the antenatal and early postnatal (T5 – 8 weeks) phase in BCHADS and then the sharp increase in mean scores from 6 months postnatal onwards, it was decided to apply the threshold validated in the antenatal period to the 8 week BCHADS assessments. As a result, the present study validates two cut-offs, one for the immediate perinatal period, and one for the later postnatal period going through into infancy.

Table 2: Summary of EPDS scores showing means, SDs, and number of cases.

Assessment	N	Mean (SD)	Total Cases
Cut-off = ≥3			
Antenatal Validation	150	2.93 (5.32)	45 (30.0%)
T1 – 1st trimester	695	2.34 (5.10)	168 (24.2%)
T2 – 2nd trimester	698	2.30 (4.56)	177 (25.4%)
T3 – 3rd trimester	607	1.73 (4.09)	110 (18.1%)
T5 – 8 weeks postnatal	545	1.79 (4.28)	105 (19.3%)
Cut-off = ≥10			
Postnatal Validation	98	7.61 (5.37)	33 (33.7%)
T6 – 6 months postnatal	415	4.34 (5.75)	64 (15.4%)
T8 – 12 months postnatal	549	4.20 (5.39)	87 (15.8%)
T9 – 24 months postnatal	674	4.17 (5.09)	97 (14.4%)

3.3. Antenatal AUC/ROC Curve Analysis

The receiver operating characteristic (ROC) curve comparing the antenatal EPDS to the MINI diagnosis of depression is presented in figure 1. The resulting area under the curve indicated that the EPDS is a fair quality instrument at this time-point (AUC = 0.734, 95% CI: 0.64, 0.80). A score of 3 or more was selected as the optimal cut-off to screen for probable perinatal depression in women (sensitivity = 62.5%, specificity = 73.9%).

3.4. Postnatal AUC/ROC Curve Analysis

The ROC curve comparing the postnatal EPDS to the MINI diagnosis of depression is presented in figure 2. The resulting area under the curve indicated that the EPDS is an excellent quality instrument at this time-point (AUC = 0.945, 95%: 0.88, 0.98). A score of 10 or more was selected as the optimal cut-off to screen for probable perinatal depression in women (sensitivity = 95.2%, specificity = 83.1%).

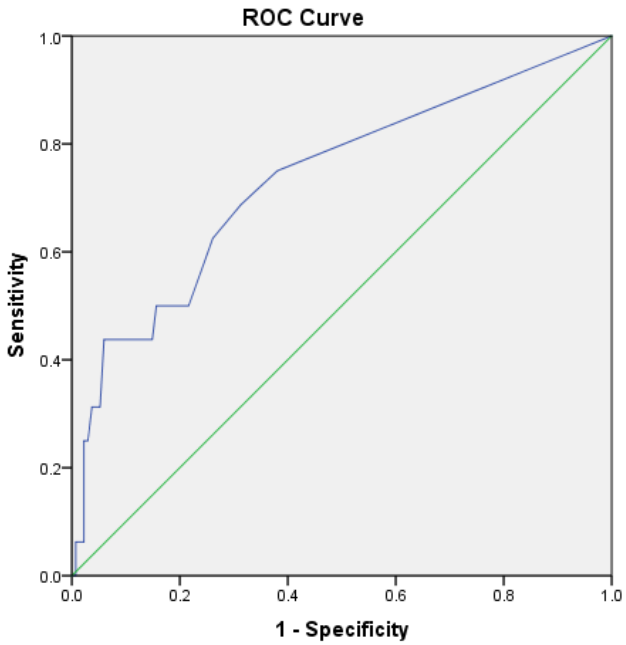


Figure 3: EPDS Receiver-operating characteristic (ROC) curve for the EPDS predicting clinically diagnosed depression during the antenatal period, sensitivity vs. 1-specificity. Area under the curve=0.734.

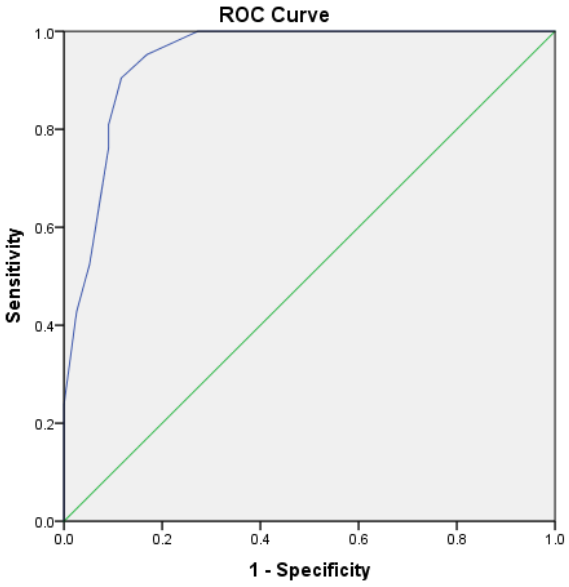


Figure 2: Receiver-operating characteristic (ROC) curve for the EPDS predicting clinically diagnosed depression during the postnatal period, sensitivity vs. 1-specificity. Area under the curve=0.95.

4. Summary of Findings

Cut-offs for detecting probable maternal depression using a Kannada version of the EPDS were validated in the independent antenatal and postnatal samples.

4.1. Antenatal Cut-Off

Results from the antenatal sample indicated that the EPDS is fair quality instrument in this period and a cut-off of 3 or more was selected for detecting probable depression (sensitivity = 62.5%, specificity = 73.9%). Mean EPDS scores during this period were quite low and a similar pattern was present in the antenatal and immediate postnatal (8 weeks) periods in the main BCHADS sample. Therefore the antenatal cut-off will be applied to the immediate perinatal period, including assessments at each trimester of pregnancy and at 8 weeks postnatal.

Using this threshold, if you surveyed 100 women in the general population (prevalence 20%), 12.5 of the 20 women with depression would be correctly identified as depressed and 20.9 women incorrectly identified as depressed. If you surveyed 100 women in a clinic sample (prevalence 50%) 31.3 of the 50 women would be correctly identified as depressed and 14.7 out of 50 women would be incorrectly identified as depressed.

4.2 Postnatal Cut-off

Results from the postnatal sample indicated that the EPDS is an excellent quality instrument in this period and a cut-off of 10 or more was selected for detecting probable depression (sensitivity = 95.2%, specificity = 83.1%). This cut-off will be applied to the late postnatal period, including assessments at 6, 12 and 24 months.

Using this threshold, if you surveyed 100 women in the general population (prevalence 20%), 19.0 of the 20 women with depression would be correctly identified as depressed and 13.5 women incorrectly identified as depressed. If you surveyed 100 women in a clinic sample (prevalence 50%) 47.6 of the 50 women would be correctly identified as depressed and 8.5 out of 50 women would be incorrectly identified as depressed.

English Version of EPDS

How have you been feeling in the past week …?

As you have recently had a baby, we would like to know how you are feeling now.
Please <u>underline</u> the answer which comes closest to how you have felt IN THE PAST WEEK, not just how you feel today.

Here is an example, already completed:
I have felt happy:
Yes, all the time
<u>Yes, most of the time</u>
No, not very often
No, not at all
This would mean "I have felt happy most of the time" during the past week. Please complete the other questions in the same way.

In the past seven days

1 **I have been able to laugh and see the funny side of things:**
As much as I always could usual
Not quite so much now
Definitely not so much now
Not at all

2 **I have looked forward with enjoyment to things:**
As much as I ever did
Rather less than I used to
Definitely less than I used to
Hardly at all

3. **I have blamed myself unnecessarily when things went wrong:**
Yes, most of the time
Yes, some of the time
Not very often
No, never

4. **I have been anxious or worried for no good reason:**
No, not at all
Hardly ever
Yes, sometimes
Yes, very often

5. **I have felt scared or panicky for no very good reason:**
Yes, quite a lot
Yes, sometimes
No, not much
No, not at all

6. **Things have been getting on top of me:**
Yes, most of the time I haven't been able to cope at all
Yes, sometimes I haven't been coping as well as usual
No, most of the time I have coped quite well
No, I have been coping as well as ever

7. **I have been so unhappy that I have had difficulty sleeping:**
Yes, most of the time
Yes, sometimes
Not very often
No, not at all

8 **I have felt sad or miserable:**
Yes, most of the time
Yes, quite often
Not very often
No, not at all

9. **I have been so unhappy that I have been crying**
Yes, most of the time
Yes, quite often
Only occasionally
No, never

10. **The thought of harming myself has occurred to me:**
Yes, quite often
Sometimes
Hardly ever
Never

EPDS - Back translation of Kannada Version

If you are pregnant or if you have just given birth to a child, we would like to know about your perspectives about a few things. Tell us about how you have been feeling, not just today, but in the past 7 days (by marking the most relevant option to the statements given below).

#1. No matter what I encounter with, I am able to find the funny aspect in that situation

0. Just how much I could do before
1. Not quite so much now
2. Definitely not as much as before
3. Not at all

#2. I look forward to (been eager to) enjoying everyday activities of life

0. As much as I used to before
1. Rather less than I used to
2. Definitely less than before
3. Hardly ever

#3. I have blamed myself when the situation went wrong.

0. No, never
1. No, once a while
2. Yes, sometimes
3. Yes, most of the times

#4. I get anxious and worried for small matters

0. Not at all
1. Once a while
2. Yes, sometimes
3. Yes, most of the times

#5. I get scared or panicky for small matters

0. Not at all
1. Not much
2. yes, sometimes
3. yes, most of the times

#6. It is becoming impossible for me to cope with stress from various tasks

0. No, I am handling it as I have done always
1. No, most of the times I have been able to manage
2. Yes, sometimes I have not been able to manage it as usual
3. Yes, most of the times it gets difficult for me to handle

#7. I find it difficult to sleep, as I have feel unhappy

0. No, never
1. No, once a while
2. Yes, sometimes
3. Yes, most of the times

#8. I have been sad or depressed

0. No, never
1. No, once a while
2. Yes, often
3. Yes, most of the time

#9. As I feel sad, I have been crying

0. Not at all
1. Once a while
2. Yes, sometimes
3. yes, quite often

#10. I have had thoughts of harming or hurting myself

0. Not at all
1. Rarely
2. Sometimes
3. Yes, most of the times

Appendix 5: NICHD Scale for Sensitivity to Non-Distress

SENSITIVITY/RESPONSIVENESS TO NONDISTRESS

This scale focuses on how the parent observes and responds to the child's social gestures, expressions, and signals as

The key defining characteristic of a sensitive interaction is that it is child-centered. The sensitive parent is tuned to the child manifests awareness of the child's needs, moods, interests, and capabilities, and allows this awareness to guide his/her interaction.

If the child initiates social gestures and expressions (looking at the parent, reaching toward the parent, waving, clapping hands, handing objects, vocalizing), or makes demands, desires, or requests known (stretching arms to be picked up, reaching for toys the parent is holding), the sensitive parent responds appropriately.

If the child loses interest, the sensitive parent takes time to re-engage the child in a manner that demonstrates sensitivity to the child's mood. When the child is bored or frustrated, the parent offers toys or other distractions. When the child is interested and involved with toys, the sensitive parent allows him/her to independently explore them. During play, the sensitive parent provides one toy or game at a time and bases continuation on the child's response. How and what they play is geared to whether or not the child seems to be enjoying the activity. The parent does not persist with an activity or toy that the child is obviously not enjoying.

A sensitive parent provides stimulation that is appropriate to the situation. He/she provides the child with contingent vocal stimulation and acknowledges the child's interest, efforts, affect, and accomplishments.

Sensitive parents can spend some time watching the child, but the difference between them and the detached parent is that the sensitive parent seems to be actively taking an interest in the child's activities, as evidenced by comments and embellishments when the child loses interest. It is at these times--when the child loses interest or is distracted--that the difference between the sensitive parent and the detached, under stimulating parent is most easily seen; the detached parent does not respond, responds in a listless manner, or responds with developmentally inappropriate comments and behavior. The insensitive parent could also be overstimulating/intrusive and might continue in his/her attempts to engage the child even when the child is providing clues that he/she is seeking to end the interaction.

A sensitive interaction is well timed and paced to the child's responses, a function of its child-centered nature. Such an interaction appears to be "in sync". The parent paces games or toy presentation to keep the child engaged and interested, but also allows him/her to disengage in order to calm down and reorganize his/her behavior. Sensitivity involves judging what is a pleasurable level of arousal for the child and helping the child to regulate arousal and affect. When the child loses interest, the sensitive parent switches

to a new tactic or toy and observes the child's reaction, or stops interacting entirely. In this way the sensitive parent can be distinguished from both an intrusive and a detached parent.

Markers of sensitivity include:
(a) acknowledging the child's affect;
(b) contingent vocalizations by the parent;
(c) facilitating the manipulation of an object or child movement;
(d) appropriate attention focusing;
(e) evidence of good timing paced to the child's interest and arousal level;
(f) slowing the pace when the child appears over stimulated or tired (e.g., demonstrates gaze aversion, fussiness);
(g) picking up on the child's interest in toys or games;
(h) shared positive affect;
(i) encouragement of the child's efforts;
(j) providing an appropriate level of stimulation when needed; and
(k) sitting on floor or low seat, at the child's level, to interact.

Thus, the sensitive parent demonstrates the ability to adapt interactions to the child's mood and level of development. The parent neither over-nor underestimates. The parent knows when it is time to increase or reduce the amount of stimulation the child is experiencing. For example, the parent discontinues an activity that is beyond the child's capacity for response or introduces a new activity when the child appears bored. Sensitive parents attend to and follow the child's lead. Ratings on this scale should be based on both quality and quantity of parent behavior.

Sensitivity/responsiveness to nondistress

1 = Not at all characteristic. There are almost no signs of parent sensitivity. Thus, the parent is either predominantly intrusive and/or detached. The parent rarely responds appropriately to the child's cues, and does not manifest an awareness of the child's needs. Interactions are characteristically ill timed or appropriate. If there is a response, it is only after the child becomes very demanding, and the response is so delayed that it cannot be construed to be contingent upon the child's behavior.

2 = Minimally characteristic. This rating should be given to parents who display infrequent or weak sensitivity/responsiveness. While the parent is sometimes sensitive, the balance is clearly in the direction of insensitivity. The parent may give some delayed perfunctory responses to cues. The parent responds rarely or slowly to the child's signals (e.g. vocalizations, affect), and appears more unresponsive than responsive. The responses tend to be minimal or perfunctory.

3 = Somewhat characteristic. This rating should be given to parents who are a mixture of sensitively responsive and not responsive to the child's signals and needs. Some

responses to signals and needs are clearly sensitive, but there are also clear signs of insensitivity in which the mother is not responsive, not fully engaged, and/or does not appropriately adjust a pace of interacting that begins to overwhelm the child. The mother can also be characterized as responsive to the child, but with behaviors that may be mechanical in quality and/or ill paced. The interaction can be characterized by a mix of well-timed and episodes that are too fast or otherwise mistimed. The parent appears to be trying to be sensitive, but has signs of insensitivity. The parent is inconsistently sensitive and may be hard to categorize.

4= Moderately characteristic. This rating should be given to parents who are predominantly sensitive/responsive. The parent demonstrated sensitivity in most interactions but may neglect to give a fuller response or a well-timed or appropriate response. The parent typically responds promptly to the child's signals, but there is some time in which clear child signals do not receive a response or in which the response is somewhat delayed or mismatched to the child's apparent need. A few of the parent's responses may be mixed, i.e. some are half-hearted or perfunctory, but the majority are full responses.

5 = Highly characteristic. This rating should be given to parents who are exceptionally sensitive and responsive. Instances of insensitivity are rare and never striking. Interactions are characteristically well timed and appropriate.

Appendix 6: Qualitative feedback from AV pilot.

Participant	Reported thought process when answering	Item Feedback
1	I felt somewhere in their positions.	NA
2	I was trying to understand how difficult it is for them	NA
3	How all can they cope?	NA
4	My family stories are also very similar. I could imagine myself in their place.	NA
5	I don't feel it's like my story but normally happens to everyone	NA
6	Nothing	1A: Didn't understand properly 1D: Didn't understand 10D: Difficult to answer
7	Nothing	8B: Didn't understand first 2 times.
8	I felt like it was my life.	6C: Finding it difficult to match options to story
9	What happens in my house everyday, also happens in their lives.	NA
10	I experienced some of stories in my life.	1C: Didn't understand middle sentence
11	While answering some questions I felt it's my sisters life stories.	NA
12	I didn't feel anything	NA
13	The characters which comes in stories I feel that it resembles my life	9D: Difficult to understand
14	I was trying to think what decisions I would take if I was in their place.	2F: Confused by the question 2E: She is anticipating mother feeling better again
15	While answering I imagine myself as a characters from some stories.	NA
16	I felt somewhere in their positions	NA
17	I was thinking how this might happen in my life also in the future.	NA
18	I was thinking of myself and answering	2D I did not understand the statement, especially the last sentence. 6F: Too long 1E: I lost track
19	I was trying to understand their situation	NA
20	I felt some stories like mine.	4B: Vignette should be re-recorded (Anandi) 1D: Little difficult to understand
21	I felt somewhere in their positions	NA
22	I felt some stories like mine, but that suicide stories I don't like.	1D: It was long
23	I felt somewhere in their situation.	4D: Too long 9D: Difficult to understand
24	I felt like this has happened in my life.	1E: Felt difficult choosing options (first AV) 6B: Did not understand sentence easily 10F: I did not listen properly

25	I felt like I had experienced the same things this week.	NA
26	I felt like all the stories happing every day in my life.	6F: Difficult to understand (specific word) 8A: Difficult to understand
27	I felt like I had to concentrate be lot	NA
28	I was thinking what to answer	NA
29	It might happen to me one day. Some have already happened.	NA
30	I did not feel anything	NA
31	I also think like them.	NA
32	All of them were going through something like me.	NA

Appendix 7: AV modifications following pilot (items 2-10)

2. I have looked forward with enjoyment to things.		
Back Translated Vignette (Changes following BT in red)	Post Pilot Modifications	Modification Rationale
A. Lakshmi feels totally hopeless about the future. She feels like she has nothing to look forward to in life and doesn't enjoy activities that she once enjoyed doing	Lakshmi feels totally hopeless about the future. She feels like she has nothing to look forward to in life and doesn't enjoy activities that she once enjoyed doing	**NO CHANGES MADE FOLLOWING PILOTING**
B. Sameena normally looks forward to happy situations but has been feeling very bad about the future lately. About once a week she is able to focus on some positive issue for a couple of hours but the rest of the time she is hopeless.	Sameena normally looks forward to happy situations but has been feeling very bad about the future lately. About once a week she is able to focus on some positive issue for a couple of hours but the rest of the time she is hopeless.	**NO CHANGES MADE FOLLOWING PILOTING**
C. Meena used to wake up with enthusiasm every day but recently she has been waking up feeling like she has nothing to look forward to. There are only a couple of times a week when she wakes up feeling good about the day ahead.	Meena used to wake up with enthusiasm every day but recently she has been waking up feeling like she has nothing to look forward to. There are only a couple of times a week when she wakes up feeling good about the day ahead.	**NO CHANGES MADE FOLLOWING PILOTING** **Although B & C a look out of place in the ranking, the mean scores suggest that the issue is that D & E are being over-rated rather than B & C being under-rated, so changes have been made to decrease intensity of D & E.**
Radha has been feeling less enthusiastic about the future recently. Every week she goes through 3 or 4 days overall when she has no interest in things that she would interest her before. D.	Every week Radha goes through 3 or 4 days where she has normal enjoyment of things. The rest of the time she has been feeling less enthusiastic about the future and has no interest in things that would normally interest her	HS and AP agreed that the focus should remain on positive feelings as this is what is cued in the EPDS item. By focusing on the negative frequency in the original vignette we may be introducing primacy effects and giving a more negative picture. Also,

		switching from descriptions of positive feelings to negative feelings is confusing and could produce inconsistent ratings.
E. Sheela enjoys life and looks forward to her future. However lately for some time she is not as enthusiastic about her future. She feels this way once a week through the day. During times like these she is unable to feel interest in activities that used to interest her.	Sheela enjoys life and looks forward to her future 6 days a week. However, lately once a week for a short period she is not enthusiastic about her future. During times like these she is unable to feel interest in activities that normally interest her.	Focus and frequency emphasis switched to the positive. Also changed **for some time** to **a short period** as the former conveys a higher level of intensity than needed for this vignette.
F. Shweta loves life and really enjoys her day to day activities. She finds that there is at least one thing every day that she looks forward to.		**NO CHANGES MADE FOLLOWING PILOTING**

3. I have blamed myself unnecessarily when things went wrong.		
Back Translated Vignette (changes made after BT in red)	Post Pilot Modifications	Modification Rationale
A. Lakshmi always feels responsible when things go wrong, even when the situation is out of her control. In a day there are normally 3 or 4 times when she will blame herself for something that was not her fault.	Lakshmi always feels responsible when things go wrong, even when the situation is out of her control. **Each** day there are normally 3 or 4 times when she will blame herself for something that was not her fault.	Changed "in a day" to "each day" to emphasise the fact that this is something that occurs every day. This needs to be in here to help distinguish this vignette from vignette C. **Name in recording may be incorrect.**
B. Sameena blames herself every day for one thing or the other although it's not her fault. Although she doesn't blame herself for everything, there is always one or the other thing that she blames herself for. Kannada vignette changed to "any one thing". PC stated that saying "one thing" would be taken too literally in Kannada and would not be equivalent in meaning to English.	**Every day** Sameena blames herself for one thing or the other although it's **definitely** not her fault. Although she doesn't blame herself for everything, there is always one thing or the other that she blames herself for.	Moved "every day" at to the beginning of the vignette to emphasise frequency. Also added "definitely" to emphasise that she is blaming herself unnecessarily.
C. When things go wrong Meena has a tendency to take the blame. There were 3 or 4 times this week when she held herself responsible for situations that she had no control over.	Recently, Meena has been blaming herself when things go wrong. There were 3 or 4 times this week when she held herself responsible for situations that she had no control over.	Vignette adapted to remove the word "tendency" as this is trait-like or dispositional in nature and so may add unnecessarily to vignette intensity
D. Radha is aware of things going on around her. Even if something is not her fault and others don't think it's her fault, she takes responsibility for it. This has happened twice a week. Kannada version changed to reflect English version.	Radha is aware of things going on around her. Even if something is not her fault and others don't think it's her fault, she takes responsibility for it. This has happened twice a week. Kannada version changed to reflect English version.	**NO CHANGES MADE FOLLOWING PILOTING**

E. Sheela sees matters the way they are and doesn't hold herself responsible for things without any reason. Once in a week she may hold herself responsible unnecessarily for something that is not her fault.	Sheela sees matters the way they are and doesn't hold herself responsible for things without any reason. Once in a week she may hold herself responsible unnecessarily for something that is not her fault.	**NO CHANGES MADE FOLLOWING PILOTING**
F. When things go wrong she doesn't hold herself responsible and blame herself. She will only blame herself when she is absolutely sure that it is her fault.	When things go wrong she doesn't hold herself responsible and blame herself. She will only blame herself when she is absolutely sure that it is her fault.	**NO CHANGES MADE FOLLOWING PILOTING**

4. I have been anxious or worried for no good reason		
Back Translated Vignette (changes after BT in red)	Post Pilot Modifications	Modifications made by PC and MBD
A. Lakshmi always worries about her health. She is in good health and there is no real reason for her worry, but there are 3 or 4 times a day where she is so worried about her health that she can't concentrate on anything else.	Lakshmi always worries about her health. She is in good health and there is no real reason for her worry, but there are 3 or 4 times a day where she is so worried about her health that she can't concentrate on anything else.	**NO CHANGES MADE FOLLOWING PILOTING**
B. Sameena wakes up once every night with anxiety and fear about her life. It is very severe at the time but in the morning she feels a lot better but doesn't understand why she felt so	Sameena wakes up once every night with anxiety and fear about her life. It is very severe at the time but in the morning she feels a lot better but doesn't understand why she felt so	**NO CHANGES MADE FOLLOWING PILOTING**
C. Meena's life seems to be going fine. She generally enjoys her day-to-day activities but suffers from "crippling" anxiety 3 or 4 times a week. These last for about an hour. She doesn't know when and how they occur but finds it impossible to carry on with her daily tasks until they stop. Kannada version changed to reflect severity in English vignette.	Meena's life seems to be going fine. She generally enjoys her day-to-day activities but suffers from **anxiety** 3 or 4 times a week. These last for about an hour. She doesn't know when and how they occur but finds it impossible to carry on with her daily tasks until they stop.	"Crippling" removed as this could be raising the vignette to a higher ranking.
D. Radha is a confident woman who enjoys spending time with her children and friends. However, twice a week without any reason she experiences severe anxiety. She can't say where and how these feelings occur and finds it difficult to focus on anything else for about half an hour until the anxiety goes away	Radha is a confident woman who enjoys spending time with her children and friends. However, twice a week without any reason she experiences severe anxiety. She can't say where and how these feelings occur and finds it difficult to focus on anything else for about half an hour until the anxiety goes away	**NO CHANGES MADE FOLLOWING PILOTING**
Sheela is always happy and enthusiastic about her life. She enjoys her work and her married life. She feels worried once a week although there is no clear reason for her worry. If she gets involved in	Sheela is always happy and enthusiastic about her life. She enjoys her work and her married life. She feels worried once a week although there is no clear reason for her worry. If she gets involved in other activities she may be able to overcome	**NO CHANGES MADE FOLLOWING PILOTING**

other activities she may be able to overcome this feeling quickly. Kannada version changed to reflect "is able" not "may". E. English version changed to make translation easier.	this feeling quickly. Kannada version changed to reflect "is able" not "may".	
F. Shweta loves her life and doesn't worry about anything unnecessarily. She is very relaxed and calm and is able to handle most things effortlessly.	Shweta loves her life and doesn't worry about anything unnecessarily. She is very relaxed and calm and is able to handle most things effortlessly.	**NO CHANGES MADE FOLLOWING PILOTING**

5. I have felt scared or panicky for no very good reason		
Back Translated Vignette (changes after BT in red)	Post Pilot Modifications	Modification Rationale
A. Lakshmi is a very tense about something untoward happening and feels constantly anxious. There are 3 or 4 times a day when she feels so fearful that she can barely move/walk. There is no specific trigger for these feelings neither is there any warning about when she will feel this way No good translation for "on edge" so changed to "anxious" in source vignette.	Lakshmi is a very tense about something untoward happening and feels constantly anxious. There are 3 or 4 times a day when she feels so fearful that she can barely move/walk. There is no specific trigger for these feelings neither is there any warning about when she will feel this way	**NO CHANGES MADE FOLLOWING PILOTING** **Although this is out of position in the rankings it is only out by one rank and it was thought that attempting to change the vignette with no obvious rationale was not worth it on balance of work created and potential outcomes.**
B. Sameena is anxious and feels fearful/ anxious through the day. She doesn't know why she feels so. Once a day the feeling builds up to a severe level of anxiety that makes it difficult to continue whatever she was doing.	Sameena is anxious and feels fearful/ anxious through the day. She doesn't know why she feels so. Once a day the feeling builds up to a severe level of **anxiety** that makes it difficult to continue whatever she was doing.	It would be better if we can change to "panic" as this is what is stated in EPDS item and English vignette. **Is there no word for panic in Kananda?** We will need to be able to explain **why** the vignettes do not reflect panic when defending theses, so need to find out if the word is there or not.

		MBD has spoken to PC who says there is no direct translation of the word "panic". The closest words translate to anxiety or heightened fear. Both 5B and 5D have used the word for anxious which I believe is more appropriate than heightened fear.
C. Meena enjoys her life. 3 or 4 times a week she suddenly feels very fearful. This can happen at any time. When this happens, she is unable to continue what she is doing, or go back to sleep, for about an hour	Meena enjoys her life. 3 or 4 times a week she suddenly feels very fearful. This can happen at any time. When this happens, she is unable to continue what she is doing, or go back to sleep, for about an hour	**NO CHANGES MADE FOLLOWING PILOTING**
D. Radha lives a peaceful life and has no reason to feel anxious or scared. Most of the time this is the case and she is able to live happily. However, twice a week she feels very anxious and frightened for about half an hour.	Radha lives a peaceful life and has no reason to feel anxious or scared. Most of the time this is the case and she is able to live happily. However, twice a week she feels very **anxious** and frightened for about half an hour.	English vignettes says "panicky"
E. Sheela feels a sense of security and happiness in her family. She is not scared about her future. Once a week she feels very fearful and has to stop what she is doing to calm herself down.	Sheela feels a sense of security and happiness in her family. She is not scared about her future. Once a week she feels **fearful** and has to stop what she is doing **for 20 minutes** to calm herself down.	"Very" removed to reduce intensity and "for 20 minutes" added to help distinguish between this and other vignettes which reference time period.
F. Shweta doesn't get anxious easily. She is a calm individual and she does not anxious without good reason. She is a source of support for her family. She brings her family together with her calm and confident demeanour.		**NO CHANGES MADE FOLLOWING PILOTING**

6. Things have been getting on top of me.		
Back Translated Vignette (Changes after BT in red)	Post Pilot Modifications	Modifications made by PC and MBD
Lakshmi feels like her life has had ups and downs and gone out of control. The stress of her	Lakshmi feels like her life has gone out of control. The stress of her husband's illness and another child has resulted in her	**NO CHANGES MADE FOLLOWING PILOTING**

husband's illness and another child has resulted in her being tired. She feels that she is drowning in trouble/difficulty bout 3 or 4 times a day. "ups and downs" has been removed. A. "Tired" has been changed to "worn out/exhausted" and "too much to handle/cope with" has been added.	being worn out/exhausted and has become too much for her to cope with. She feels like she is drowning in difficulty about 3 or 4 times a day.	
B. Sameena has recently begun to feel burdened by her family responsibilities. This feeling grows from the time she wakes up. Once everyday she feels helpless and feels it is impossible to continue what she is doing.	Sameena has recently begun to feel **over**burdened by her family responsibilities. This feeling grows from the time she wakes up. Once everyday she feels helpless and feels it is impossible to continue what she is doing.	Burdened has been changed to "over burdened" to match intensity described in C. **Vignettes in this set are only one rank out of place so only require small changes.**
Meena is normally able to manage her work but there have been 3 or 4 times this week when she has felt burdened by her work. The work she has to do hasn't changed but it still feels like it has increased. C. Changed to "extremely burdened" to reflect severity of English version.	Meena is normally able to manage her work but there have been 3 or 4 times this week when she has felt extremely burdened by her work. The work she has to do hasn't changed but it still feels like it has increased.	**NO CHANGES MADE FOLLOWING PILOTING**
D. Radha was always able to effectively manage all her difficulties. However, lately, couple of times a week she felt very stressed about handling these household responsibilities.	Radha was always able to effectively manage all her difficulties. However, lately, couple of times a week she felt very stressed about handling **all** these household responsibilities.	"all" added to increase feeling of being overwhelmed my responsibilities, too much to handle etc.
E. Sheela has been managing her work/responsibilities well this week. On one occasion she felt as if she wasn't able to. She used to take care of her sister's children once a week in the past but off late found it difficult to do so. Kannada version changed to say that taking care of children is continuous activity, not in the past.	Sheela has been managing her work/responsibilities well this week. On one occasion she felt as if she wasn't able to. She used to take care of her sister's children once a week in the past but **this week** found it difficult to do so. Kannada version changed to say that taking care of children is continuous activity, not in the past.	Changed "of late" to "this week" to minimise time frame of difficulties and therefore reduce intensity of vignette.

Shweta is a woman in control. She is very self-sufficient and has is capable of handling anything. She feels proud about her ability to handle stress/pressure and hates asking for help from anyone. The day thing can go (VS). F. English version edited.	Shweta is a woman in control. She is very self-sufficient and has is capable of handling anything. She feels proud about her ability to handle stress/pressure and hates asking for help from anyone.	**NO CHANGES MADE FOLLOWING PILOTING**

7. I have been so unhappy that I have had difficulty sleeping.		
Back Translated Vignette (changes after BT in red)	Post Pilot Modifications	Modification Rationale
A. Lakshmi has been so sad lately that she has only been able to get a few hours' of sleep at night. She wakes up 3 or 4 times every night feeling hopeless and upset and takes an hour fall asleep again.	Sameena has been feeling very upset recently. This feeling is more at night when all her daily chores have ended. She gets negative thoughts and lies in bed awake for several hours every night before falling asleep.	HS and AP agreed that items A & B should be switched around in the rankings. Rationale is that taking a long time to fall asleep is likely to be seen as more of an issue than waking up several times during the night. **No changes made to text**
B. Sameena has been feeling very upset recently. This feeling is more at night when all her daily chores have ended. She gets negative thoughts and lies in bed awake for several hours every night before falling asleep. Negative does not translate well into Kannada. PC suggests asking VS for kannada word for pessimism. Says they must use it in clinical psychology.	Lakshmi has been so sad lately that she has only been able to get a few hours' of sleep at night. She wakes up 3 or 4 times every night feeling hopeless and upset and takes an hour fall asleep again.	Items A & B switched around in rankings. **No changes made to text.**
C. Half of the time, Meena sleeps right through the night, but on 3 or 4 nights of the week she is feeling so upset that she cannot sleep. On such nights it takes her about an hour to fall asleep.		**NO CHANGES MADE FOLLOWING PILOTING**
D. Radha is very dissatisfied with her work. She only works two days a week, but the thought of going to work the next day makes her restless during sleep the night before. She		**NO CHANGES MADE FOLLOWING PILOTING**

therefore does not feel rested and fresh when she wakes up.		
E. Sheela usually sleeps well at night but once a week she struggles to sleep. On those nights she feels very depressed and tends to worry about things that have gone wrong in her life and is unable to stop herself from thinking so.		**NO CHANGES MADE FOLLOWING PILOTING**
F. Shweta falls asleep within 10 minutes every night. She sleeps deep and does not wake up until her alarm goes off. She wakes up, feeling no fatigue and refreshed.		**NO CHANGES MADE FOLLOWING PILOTING**

8. I have felt sad or miserable		
Back Translated Vignette (changes after BT in red)	Post Pilot Modifications	Modification Rationale
A. Lakshmi is so sad that she tries to stay away from people whenever possible. She feels like being by herself 3 or 4 times a day. No one is able to cheer her up.	Lakshmi is so sad that she tries to stay away from people whenever possible. She feels like being by herself 3 or 4 times a day. No one is able to cheer her up.	**NO CHANGES MADE FOLLOWING PILOTING**
Sameena feels very sad in the mornings. She is so sad that she manages to get herself out of bed. But is subsequently able to do her daily chores and be with people and feel better through the rest of the day. Added "when she wakes up" after "mornings. Changed to "manages with difficulty". B. Final sentence refers immersing herself in work and people in order to forget sadness, so equivalence is preserved.	Sameena is so sad that she **struggles** to get herself out of bed **every morning**. But is subsequently able to do her daily chores and be with people and feel better through the rest of the day.	This vignette is being grossly underrated so "manages to get out of bed with difficulty" has been changed to "struggles" and "every morning" has been added to the end of the sentence, to emphasise that it is a daily difficulty. I have also removed the first sentence as it restricts feelings of sadness to mornings and it could be interpreted as fairly normal to feel sad when you wake up and have to get out of bed!!
C. Meena enjoys spending time with her family and enjoys. She is also hopeful about her future. She feels very sad and depressed 3 or 4 times a week. This has made her to lose	Meena enjoys spending time with her family and enjoys. She is also hopeful about her future. She feels very sad and	**NO CHANGES MADE FOLLOWING PILOTING**

interest in things she would enjoy doing earlier. Makes sense in Kannada.	depressed 3 or 4 times a week. This has made her to lose interest in things she would enjoy doing earlier	
D. Radha is usually known to be a happy and enthusiastic person. Couple of times a week she feels very unhappy and dissatisfied. During times like these, she prefers to be alone. She is able to overcome this and focus on something enjoyable only after about an hour.	Radha is usually known to be a happy and enthusiastic person. Couple of times a week she feels very unhappy and dissatisfied. During times like these, she prefers to be alone. She is able to overcome this and focus on something enjoyable only after about an hour	**NO CHANGES MADE FOLLOWING PILOTING**
E. Sheela is a happy person and likes to spend time with her family. She spends time with friends twice or thrice a week. However once a week for about half an hour she feels very sad but doesn't notice/view this as being significant.	Sheela is a happy person and likes to spend time with her family. She spends time with friends twice or thrice a week. However once a week for about half an hour she feels very sad but doesn't notice/view this as being significant.	**NO CHANGES MADE FOLLOWING PILOTING**
F. Shweta loves her life and is always happy. She enjoys spending time with her children. She also finds her work very rewarding.	Shweta loves her life and is always happy. She enjoys spending time with her children. She also finds her work very rewarding.	**NO CHANGES MADE FOLLOWING PILOTING**
9. I have been so unhappy that I have been crying		
Back Translated Vignette (changes after BT in red)	Post Pilot Modifications	Modification Rationale
A. Lakshmi feels so miserable that she tries to stay away from people whenever possible. She feels so sad that she cries 3 or 4 times a day.	Lakshmi feels so miserable that she tries to stay away from people whenever possible. She feels so sad that she cries 3 or 4 times a day.	**NO CHANGES MADE FOLLOWING PILOTING**
B. Sameena is finding life extremely difficult at present. She keeps herself busy and gets throughout the day but at night when she doesn't have work, she feels very sad and cries for about an hour before she falls asleep	Sameena is finding life extremely difficult at present. She keeps herself busy and gets throughout the day but at night when she doesn't have work, she feels very sad and cries for about an hour before she falls asleep **each night.**	Added "each night" to emphasise that it is something that happens every day and increase vignette intensity to put it above C.
C. Meena has been feeling really overwhelmed by situations recently. She feels a growing distance with her husband and is finding her work extremely stressful. She finds it hard	Meena has been feeling really overwhelmed by situations recently. She feels a growing distance with her husband and is finding her work extremely stressful. She finds it hard to manage this and cries 3 or 4 times a week.	**NO CHANGES MADE FOLLOWING PILOTING**

	to manage this and cries 3 or 4 times a week.		
D.	Radha enjoys life the most when she is active and busy. When she is involved in her daily activities, she hardly gets affected by ups and downs. When all of these activities end, she feels very sad, and has cried twice in a week.	Radha enjoys life the most when she is active and busy. When she is involved in her daily activities, she hardly gets affected by ups and downs. When all of these activities end, she feels very sad, and has cried twice in a week.	**NO CHANGES MADE FOLLOWING PILOTING**
E.	She loves life and is a happy person. She likes spending time with people. She however feels sad at times and has cried once a week.	She loves life and is a happy person. She likes spending time with people. She however feels sad at times and has cried once a week.	**NO CHANGES MADE FOLLOWING PILOTING**
F.	Shweta feels very happy and satisfied with her life. She sometimes feels a little sad or dissatisfied but it is not a reason for her to cry.	Shweta feels very happy and satisfied with her life. She sometimes feels a little sad or dissatisfied but it is not a reason for her to cry.	**NO CHANGES MADE FOLLOWING PILOTING**

10. The thought of harming myself has occurred to me		
Back Translated Vignette (changes after BT in red)	Post Pilot Modifications	Modification Rationale
Lakshmi is currently under unbearable amount of stress. She worries a lot about money and has recently found out that she is pregnant. She is very anxious about the future and thinks about harming herself 3 or 4 times a day ~~as a way of relieving the pressure she is feeling. She also believes that by doing so her stress will be relieved.~~ Reference to relieving pressure has been removed as does not reflect suicidality.	Lakshmi is currently under unbearable amount of stress. She worries a lot about money and has recently found out that she is pregnant. She is very anxious about the future and thinks about harming herself 3 or 4 times a day	**NO CHANGES MADE FOLLOWING PILOTING**
A. Sameena's husband died a few months ago. Since then she has been very sad and also stressed about managing her household responsibilities alone. She finds it difficult to talk to others about how she is feeling. Once a day she thinks about harming herself ~~just so that she can express what she is feeling.~~ Reference to expressing self removed as does not reflect suicidality.	Sameena's husband died a few months ago. Since then she has been very sad and also stressed about managing her household responsibilities alone. She finds it difficult to talk to others about how she is feeling. Once a day she thinks about harming herself	**NO CHANGES MADE FOLLOWING PILOTING**
B. Meena has fallen ill recently and is unable to enjoy doing things that she did earlier with joy. She is unable to help out with the children or go to work. She feels worthless as a result. There are 3 or 4 times a week where she has thoughts of harming herself ~~in order to overcome the numbness that she experiences in life.~~ Reference to numbness removed to reflect suicidality.	Meena has fallen ill recently and is unable to enjoy doing things that she did earlier with joy. She is unable to help out with the children or go to work. She feels worthless as a result. There are 3 or 4 times a week where she has thoughts of harming herself	**NO CHANGES MADE FOLLOWING PILOTING**

C.	Radha feels like she is not in control of her life. She feels as if she is experiencing one stressful situation after another. She feels as if she wants to regain control of her life and this thought crosses her mind twice a week. She thinks that the only way to do this may be to harm herself. Reference to regaining control removed to reflect suicidality.	Radha feels like she is not in control of her life. She feels as if she is experiencing one stressful situation after another. It crosses her mind twice a week to harm herself.	**NO CHANGES MADE FOLLOWING PILOTING**
D.	Sheela's family members criticize/harass her and consider her to be lazy and accuse her of not doing any work in the house like her other family members. She feels that they constantly criticize her and last week it went to an extent where she even thought of harming herself.	Sheela's family members criticize/harass her and consider her to be lazy and accuse her of not doing any work in the house like her other family members. She feels that they constantly criticize her and last week it went to an extent where she even thought of harming herself.	**NO CHANGES MADE FOLLOWING PILOTING**
E.	Shweta is under a lot of pressure at work and it has been very difficult for her to manage it. Fortunately, in the evenings she has been able to share her feelings with her family and has not gone to the point where she has thought of harming herself.	Shweta is under a lot of pressure at work and it has been very difficult for her to manage it. Fortunately, in the evenings she has been able to share her feelings with her family and has not gone to the point where she has thought of harming herself.	**NO CHANGES MADE FOLLOWING PILOTING**

Appendix 8: Anchoring Vignettes

Item1 – I have been able to laugh and see the funny side of things…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X is usually very cheerful but recently she has not been feeling like herself. Situations that normally wouldn't bother her have been putting her in a bad mood. She used to laugh every day, but she can't remember finding anything funny in the last week.	Lakshmi saadharanavaagi thumbaa khushiyaagi irutthaare. Aadhare munchinanthe anisuthilla. Maamuliyaagi avarige thondhare kodadha sandharbhagalu, eega avarannu ketta manasthithige tharutthidhe. Avaru dinaa nagunagutthaa irutthiddaru.aadhare, kaledha ondhu vaaradindha avarige yenu thamashe thandhu koduva haage jnyaapaka barutthilla.	Lakshmi will usually be very happy. But she's not feeling like before. The contexts/ situations which usually won't bother her are creating bad moods. She used be cheerful every day. But, From past one week, she can't remember things which give her joy (she can't remember any that will bring her fun).
B	X is usually very happy but work has been getting her down recently and she has been much more serious than usual. In the last week she has probably only laughed once.	Sameena saamaanyavaagi bahala santhoshadinda iruttare. Adare itthichege eladarinda kuggiddale haagu ghambeeravaagiddale. Kaleda ondu vaaradalli bahusha avaru onde baari nakkirabahudu.	Sameena will usually be very happy. But, she is in low spirits because of day to day challenges and is very serious now a day. Probably, she must have laughed only once this past week.
C	X has always been a happy and positive person, but recently she has been feeling down about life. She would normally joke around and laugh with her friends every day, but she has only laughed a couple of times in the last week.	Meena yaavagalu santhoshavaagi iruttaare. Aadhare itthichege avarige jeevanadalli kuggutthiruvanthe anubhava aagutthide. Avaru saamanyavaagi thamaashe maaduttha, yellarondhige prathidina nagunaguthaa iddharu, aadhare kaledha vaaradalli avaru kevala vondheradu baari nakkiddhaare.	Meena is happy all the time. But she's feeling low in life now a day. She usually used to make fun, used to be cheerful with everyone every day. But, she has laughed/ smiled one or two times in the last week.
D	X has been able to enjoy time with friends and family 3-4 days a week. The other half of the week she feels very low and barely able to force a smile.	Sitage thaanu munche iruvantheye anisuthade, haagu varadalli 3 athava 4 dinagalu, snehitharodane-kutumba davarodane anandisalu saadhyavaagutide. Ulida dinagalu avalu dukhadinda iddalu mattu nagalu saha kashta vagithu.	Sita feels that she has been the same as ever before, and 3 or 4 times in a week, she can enjoy the time with the friends and the family members. She was sad and had difficulty even to smile rest of the days.
E	X has been enjoying life just as she usually does for most of the week. There have been a couple of times where she has struggled to cope when things haven't gone to plan but otherwise she has been able to laugh off any difficulties.	Sheela Vaaradalli hechhu dinagalu santhoshavagi iddaare. Andhukondanthe kelasagalu nadeyadiddaaga adannu nibhayisikondu hogalu kelavomme kashta vaagutthittu. Adannu bittare kashtadallu avalu nagalu sadyavaayithu.	Sheela is happy most of the days in a week. At times had difficulty to manage when the work didn't happen in the expected way. Apart from that she could smile even in tough times.
F	X is normally very happy in life and this week has been no different. While she has been busy with different jobs that needed doing, she has been able to stop and spend some time laughing and having fun with her children 3 or 4 times a day.	Shwetha saamaanyavaagi santoshavaagi iruthare. Ee vaara kooda haage iddaru. Avaru maada bekaagiruva halavaaru kelasa karyagalunnu maadutta, kelavu samayavannu makkala jothege 3 rinda 4 baari moju mastiyalli kaleyuvaru.	Shwetha will usually be happy. She was feeling the same this week also. Along with daily chores which she is supposed to do, she spends time to have fun with children 3 or 4 times a day.

Item 2: I have looked forward with enjoyment to things			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X feels completely hopeless about the future. She never feels like she has anything to look forward to and doesn't enjoy activities that she used to love	Sowmya avaru bhavishyada bagge sampoornavaagi hathaasharaagiddhaare. Avarige jeevanadalli yeduru nodalu yenuu illadanthe anisuthade matthu avaru ee hindhe maaduthiddha chatuvatikegalu eega avarige khushi kodutthilla.	Lakshmi feels totally hopeless about the future. She feels like she has nothing to look forward to in life and doesn't enjoy activities that she once enjoyed doing
B	X normally feels like she has plenty to look forward to but has been feeling very bad about the future lately. About once a week she will manage to focus on something positive for a couple of hours but the rest of the time she feels hopeless.	Mallika avaru saadhaarnavaagi khushi sangathigalannu yedhuru noduthare, aadare itthichege avarige bhavishayada bagge tumba besaravenisuttide. Vaaradalli ondu baari avarige yaavudaadaru olleyya vishayada bagge vonderadu gantegala kaala gamana harisalu saadyavaagutthade, aadare ulida samaya avaru niraase inda iruthare.	Sameena normally looks forward to happy situations but has been feeling very bad about the future lately. About once a week she is able to focus on some positive issue for a couple of hours but the rest of the time she is hopeless.
C	X used to wake up every day excited for what was in store but recently she has been waking up feeling like she has nothing to get out of bed for. There are only a couple of times a week when she wakes up feeling good about the day ahead.	Hema pratidina yeddeluvaga uthsaahadindha aa dinavannu yedhuru noduthiddalu, aadhare itheechege aakege yelalikke yenu kaarana illavendhu anisutthide. Vaaradalli kevala ondheradu sala maathra avalu yedheluvaaga olleya bhaavane indha aa dinavannu edhuru noduthaale.	Meena used to wake up with enthusiasm every day but recently she has been waking up feeling like she has nothing to look forward to. There are only a couple of times a week when she wakes up feeling good about the day ahead.
D	X feels happy with life for 3 or 4 days a week. The rest of the time she feels discouraged and less motivated about the future. At these times she has no interest in the things she would normally look forward to.	Prathi vara Bhagya ,3 athava 4 dinagalu samanyavagi anadadinda iruthare. Ulida samayadalli avala bhavishyada bagge kammi uthsahadinda iddale hagu samanyavagi asakthi idda vishayagalalli,iga asakthi kaledu kondidale.	Bhagya usually feels happy for 3 or 4 days in a week. She feels less excited about her future for rest of the times and she has lost interest now in the activities where she used to be interested in.
E	X enjoys her life and feels like she has plenty to look forward to. However, once a week, she does not feel motivated about the future. This feeling lasts for one day. During this time she has no interest in things that she would normally be excited about.	Sumitra varadalli 6 dinagalu thanna jeevanavannu anandisuthale haagu avalu bhavishyada bagge yeduru noduthalle. Adare varadalli ondu dina swalpa samayakke avalige bhavishyada bagge uthsahaviruvudilla. Inthaha samayagalalli avalige asakthi iruthida chatuvatikegalalli ,asakthiye illadanthe agide.	Sumitra enjoys her life 6 days in a week and looks forward to the future. She doesn't feel excited about the future once in a week for some time. During then, she won't be interested in things which normally excite her.
F	X loves life and really enjoys her day to day activities. She finds that there is at least one thing every day that she is looking forward to.	Ramya jeevanavannu preethisuthale matthu nijaavagiyoo thanna dina nithayada chatuvatikegalannu anandisuthare. Kanishta ondu chatuvatikegagi prathi dina yeduru noduthiddale.	Shweta loves life and really enjoys her day to day activities. She finds that there is at least one thing every day that she looks forward to.

Item 3 – I Have blamed myself unnecessarily when things went wrong…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X always feels guilty when things go wrong, even when it is out of her control. There are normally 3 or 4 times a day where she will blame herself for something that was not her fault.	Pavithra avaru enaadharu thappaadaaga, paristhithi avara hidithadalli illaddhidharu kooda, avarannu dhooshisikolluthaare. Prathi dina moornalku baari, avara thappu illadhidharu kooda, Pavithra avaru thammanu thaavu dooshisikollutthaare.	Pavithra blames herself when mistake happens even if the situation wasn't under her control. Pavithra blames herself 3 -4 times every day for no mistakes of hers.
B	X blames herself for one thing that was not her fault every day. She doesn't blame herself for everything, but there is always that one thing.	Prathi dina Ashwiniravaru avara thappe illadidaru, yavudadaru ondu vishayada bagge avaranne avaru dooshisi kolluthare. Avaru yella vishayada bagge avarannu dooshisikolluvudilla, aadare yaavudadaru ondu vishayada bagge dooshisikolluthare.	Its a daily habit of Ashwini to blame herself for any random mistake which need not be committed by her. This blaming of her is limited to any one mistake.
C	X has a tendency to take the blame when things go wrong. There were 3 or 4 times this week when she felt responsible for accidents or situations that she had no control over.	Ittichege, enadaru tappadaga Ruksar avaru thammannu thaave dhooshisi kolluthare. Kaledha vaaradalli moorarindha naalku baari paristhithigalu avara niyanthrana dalli illadidharu, avaranne hone gaararaagi maadikondidhaare.	Nowadays Ruksar has made herself scapegoat for every mistake. She took responsibility even when the situation wasn't under her control 3 -4 times in the last week.
D	X generally has a good grasp of what's going on around her. However, every now and then she will take the blame for something that wasn't her fault, even when no one else thinks it is her fault. Recently, this has been happening around twice a week.	Gangalige thanna suttha muttha nadeyuthiruvudhara bagge arivu iruthadhe. Adhagiyu, avala thappu iradhidharu, bereyavaru saha avala thappu illa endhu bhaavisidharoo, ommomme thammannu thaave dhooshisikolluthaare. Idhu vaaradalli sumaru yeradu baari heegaaguvudhu.	Ganga is usually aware of the happenings around her. Sometimes, she blames it upon herself even though her innocence is known to the rest. Recently, this has been happening twice in a week.
E	X sees things as they are and does not make a habit of blaming herself when something is not her fault. Only once a week she will take the blame unnecessarily when something goes wrong.	Divya vishayagalu hege irutho haage noduthaale, mathu thanannu thanu vinaakaarana dooshisikolluva abhyaasavilla. Idu bahusha vaarakomme mathra avalu anagathyavagi yenadaru thappadaga thanannu thane dhooshisikolluthale.	Sheela sees matters the way they are and doesn't hold herself responsible for things without any reason. Once in a week she may hold herself responsible unnecessarily for something that is not her fault.
F	X does not feel guilty when things go wrong and she is not to blame. She will only blame herself when she is absolutely sure that she is at fault.	Rashmi avaru Vishayagalu thappagi hodaga thannannu thappikasthe yendu bhaavisuvudilla mathu avalannu dhooshisuvudilla. Avalige thappu maadiddene yendu sampoornavaagi khachitha vaadaga mathra thannannu dhooshisi kolluthale.	When things go wrong she doesn't hold herself responsible and blame herself. She will only blame herself when she is absolutely sure that it is her fault.

Item 4: I have been anxious or worried for no good reason…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X worries constantly about her health. She is in good health and there is no obvious reason for her anxiety, but there are 3 or 4 times a day where she is so worried about it that she can't concentrate on anything else.	Bhagya avaru sadaa thamma aarogyada bagge chinthe maadutthaare. Avaru arogyavaagidhaare, haagu avara aathankakke yaava sooktha kaaranavilla. Aadharu avaru dhinadalli moornaalku baari thamma aarogyadha bagge yeshtu chinthe maadutthaare yenadhare avarige bere yaavudhara melu gamana kodalu saadhyavaaguthilla.	Lakshmi always worries about her health. She is in good health and there is no real reason for her worry, but there are 3 or 4 times a day where she is so worried about her health that she can't concentrate on anything else.
B	X wakes up once every night with a great sense of anxiety and fear about her life. It is overwhelming at the time but in the morning she feels a lot better and doesn't understand why she got so worked up.	Prathi raatri ondubaari Anandiravaru avara jeevanada bagge bhaya haagu athankadinda yechharavaaguthare. Yecchara vaadhaga ee bhaavane galannu sahisalaaguvudilla. Aadare belligge avarige aaramavenisuthade haagu ee reethi anubhava yaake aayithu yendu artha aaguvadilla.	Sameena wakes up once every night with anxiety and fear about her life. It is very severe at the time but in the morning she feels a lot better but doesn't understand why she felt so
C	X's life seems to be going well. She is generally happy and enjoys her day-to-day activities but suffers from episodes of crippling anxiety a 3 or 4 times a week. These last for about an hour. She doesn't know when they are going to happen or where they come from but finds that she cannot carry on with her daily tasks until they stop.	Priya la jeevana chennagi saaguthidhe. Ake saamaanyavaagi thanna dhina nithyada chatuvatikegalannu anandisutthidalu. Aadhare vaaradalli moorarindha naalku baari aathankakke olagaaguthaare. Ee athanka sumaaru ondhu gante gala kaala iruthadhe. Idhu yaavaga? Hege? Agutthe antha gothilla. Idhu nilluva varegu yaavudhe dhina nithyadha kelasa maadalu sadhyavaaguthilla.	Priya's life going on smoothly. She usually used to enjoy her daily activities. But she feels anxious 3 - 4 times in a week. This anxiety remains for an hour. She isn't aware when? how? this happens. She can't do any daily activities until this stops.
D	X is a confident woman who enjoys spending time with her children and catching up with her friends. However, twice a week a great sense of anxiety about her life will hit her out of nowhere. She can't say where these feelings come from and finds it difficult to focus on anything else for about half an hour until they pass.	Ramya aathmavishwaasa hondhiruva vyakthi. Avaru thanna makkalodane haagu snehitharodane kaleyuva samayavannu anandhisuthare. Adagiyu, vaarakke yeradu baari jeevanada bagge vinaakarana avalige athi hecchu aathankavaaguthadhe. Ee reethi bhaavanegalu yellindha haagu hege barutthide endhu avalige helalu saadhyavaguthilla, matthu sumaaru ardha gante aa bhaya hoguvavaregu yaavudhara baggeyu gamana harisalu saadhyavaaguvudilla.	Radha is a confident woman who enjoys spending time with her children and friends. However, twice a week without any reason she experiences severe anxiety. She can't say where and how these feelings occur and finds it difficult to focus on anything else for about half an hour until the anxiety goes away
E	X is always happy and very positive about life. She enjoys her job and is happily married. She gets worried once a week, even though there is no reason for this anxiety. She is able to move past it quickly if she busies herself with other things.	Kusum yaavagalu khushiyaagi iruthale mathu jeevanada bagge utthsaahadinda iddale. Kelasavannu mathu daampanthya jeevana anandadisuthaale. Vaaradalli omme aathankakke ollagaguthaale aadare ee chinthe ge yaavude spashtavada kaaranavilla. Avalu bere kaaryagalalli niratharaagiddaga idarinda bega hora baralu saadhyavaaguthade.	Kusum is always be happy and is excited about life. She enjoys the work and marital life. She feels anxious without any specific reason once in a week. She can come out of this easily when she's involved in other activities.
F	X loves her life and doesn't worry about anything unnecessarily. She is very laid back and is able to take most things in her stride without letting them affect her mood.	Pavana thanna jeevanavannu preethisuthale mathu anagathyavaagi yennanu chinthisuvudilla. Avalu tumba aaramavaagiruthale mathu avala manasthithige paarinaama beeladhanthe tumba vishayagalannu nibhayisuthale.	Shweta loves her life and doesn't worry about anything unnecessarily. She is very relaxed and calm and is able to handle most things effortlessly.

Item 5: I have felt scared or panicky for no good reason…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X is a very jumpy individual and feels constantly anxious. There are 3 or 4 times a day when she will feel absolutely frozen with fear. There is no specific trigger for these feelings, so there is often no warning about when they will strike.	Selvi avaru yaavaga yenu aagabahudu yendu bhaya dhindha irutthaare,Sadaa chadapadisuvanthe anubhavavaaguthhithu. Avaru dhinadalli moornaalku baari yeshtu bhayadhindha irutthare yendhare chalisalu kooda kashtavenisutthadhe. Ee anubhavakke yaava sooktha kaaranavilladhidharu ee anubhava yaava munsoochane illadhe agutthadhe.	Lakshmi is a very tense about something untoward happening and feels constantly anxious. There are 3 or 4 times a day when she feels so fearful that she can barely move/walk. There is no specific trigger for these feelings neither is there any warning about when she will feel this way
B	X is a nervous individual and feels a sense of anxiety throughout the day. She doesn't know where the feeling is coming from. Once a day the feeling builds up to a level of panic that makes it impossible to continue whatever she was doing at the time.	Dilshad avaru dinavidee bhaya bhayadinda haagu athankadinda chadapadisuthare. Avarige ee reethi anubhava yaake aaguthide yendu arthavaaguthilla. Dinakke ondu baari ee anubhava tumba theevra vaagi avaru ghaabariyaaguthare. Idarinda avarige maaduthida kelasavannu munduvaresalu asaadhya vaaguthide	Sameena is anxious and feels fearful/ anxious through the day. She doesn't know why she feels so. Once a day the feeling builds up to a severe level of anxiety that makes it difficult to continue whatever she was doing.
C	X enjoys her life but 3 or 4 times a week she is suddenly struck with fear. This can happen at any time and she is unable to continue what she is doing, or go back to sleep, for about an hour until the episode passes.	Neha jeevanavannu aanandhisuthaale, aadhare vaaradalli moorarindha naalku baari iddakkiddanthe athi bhaya beeluthaare. Idhu yaava samayadalladharu aagabahudhu. Ee reethi aadhaga avalu maadutthiruva kelasa-kaaryagalannu mundhuvaresalu athava nidhre maadalu saha ondhu gantegala kaala aguvudhilla.	Meena enjoys her life. 3 or 4 times a week she suddenly feels very fearful. This can happen at any time. When this happens, she is unable to continue what she is doing, or go back to sleep, for about an hour
D	X lives a peaceful life and has no reason to feel anxious or scared. Most of the time this is the case and she is able to get on with life very happily. However, twice a week she feels panicky and frightened for about half an hour.	Meghana nemmadhiyindha jeevana nadesuthidhaare matthu avalige bhaya athava aathanka paduvanthaha yaavudhe kaarana illa. Bhahalashtu samaya idhe reethi iddhu, aakege santhoshadhindha jeevisalu saadhyavaguthadhe. Aadharu vaaradalli eradu baari avalige ghaabari matthu bhayada bhaavane ardha gante aadharu anisuthadhe.	Radha lives a peaceful life and has no reason to feel anxious or scared. Most of the time this is the case and she is able to live happily. However, twice a week she feels very anxious and frightened for about half an hour.
E	X is a calm individual who feels secure and happy in her family. She does not fear the future but once a week she becomes very panicky and has to stop what she's doing in order to calm herself down.	Rekha thanna kutumbadalli surakshitha matu santoshavannu anubhavisuva vyakthi. Avalu bhavishyada bagge bhaya paduvudilla. Aadare vaarakomme bhaya bheethalaaguthale mathu thannanu samaadhaana golisikollalu avalu maaduthidda kelasavannu 20 nimishagala kala nillisuthale.	Rekha is the person who experiences safe and happiness in her family. She doesn't fear about the future. But she feels fearful once in a week and she stops the activity for 20 minutes to console herself.
F	X is not easily panicked. She is a calm individual and she does not panic or feel afraid without good reason. She is the rock of her family and often holds things together with her confident and peaceful demeanour.	Yasmeen sulabhavaagi ghaabariyaguvudilla avalu shaantha vyakthi mathu avalu olleya kaaranavillade ghaabariyaaguvudilla athava hedaruvudilla. Avalu kutumbakke aadhaaravaagidhalle. Thanna aatma vishwaasa matthu shanthiyutha varthane yinda aagaga thanna kutumbhavannu ottu goodisiddale.	Shweta doesn't get anxious easily. She is a calm individual and she does not anxious without good reason. She is a source of support for her family. She brings her family together with her calm and confident demeanour.

Item 6: Things have been getting on top of me…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X feels like her life has been spiralling out of control recently. The combined pressure of an extra child and her husband's chronic illness has worn her out and become too much for her to handle. She feels out of her depth 3 or 4 times a day.	Poornima avara jeevana avarige thumba ne hathotige sigadhanthe anisuthidhe. Gandana kaayile haagu innondhu magu, ee eradoo kaaranadhindha aagiruva otthada avarige thadeyuvudakke asaadhyavaagidhe haagu susthaadhanthe agidhe. Dhinakke sumaaru moornaalku baari avarige kashtadalli mulugidhanthe anisuthadhe.	Poornima feels her life is going out of her control. She is finding it tough to handle the stress due to her husband's illness and another baby and feels tired. She feels as though she's sunk in the problems 3 – 4 times a day.
B	X has recently begun to feel completely snowed under by her family commitments. This feeling builds up from the moment she wakes up. Once a day she gets to a point where she feels broken by it and completely unable to carry on.	Chaithra avarige itthichege kutumbada javabdaarigalinda hecchu hore yenisuthide. Avaru belligge yeddagininda ee anubhava hechuuttha hoguthade. Dinakke ondu baari ee anubhavadinda avarige sothanthe anisutthade haagu kelasagallannu munduvaresalu asaadhya vaagutthade.	Chaithra feels burdensome due to responsibilities of her family. This sense keeps increasing throughout the day since dawn. At least once a day she feels like she has failed because of this pressure and it affects her progress of work.
C	X is normally able to manage her workload but there have been 3 or 4 times this week when she has felt like it has been getting on top of her. The amount she has to do hasn't changed but it just feels like too much.	Baseera saamaanyavagi thanna kelasa kaaryagalannu nibhayisuthaare. Aadhare ee vaaradalli moorarindha naalku baari avarige ee kelasagalu thumba saakagidhe endhu anisuthidhe. Avalu maadabekaada kelasa karyagalu badhalaagilla. Aadharu adhu thumba hecchaadhanthe anisutthide.	Baseera usually handles her work. But 3-4 times this week, she has felt that work was tiresome. The work she is supposed to do hasn't changed. Despite she feels it is too much.
D	X has always been able to handle everything life throws at her. Recently though, there have been a couple of times a week when all the pressures of looking after her family have felt like too much for her to deal with.	Shahim jeevanadalli yaavagalu yene kashta edhuraadaru adannu nibhayisuthiddaru. Aadharu, itthichege vaardalli kelavu baari thamma ella mane kelasagalannu maduva javabdariya otthadagalu athyanta kashtakara enisutthidhe.	Shahim could always handle any problem in her life. But, nowadays once in a while in a week, she feels that the pressure of finishing the responsibility of all her household work is very difficult.
E	X has been handling things well this week apart from one occasion where she felt like she wasn't coping. She looks after her sister's children once a week and has always managed it in the past but found it very difficult this time round.	Shama kelasa kaaryagalunnu ee vaaradalli chennagi nibhayasuthidale. Aadare omme nibhayisalaare yemba anubhava hondiddale. Thanna akkana makkalannu vaarakke ondu baari kashtavillade nodikolluthidalle. Aadare ee vaaradalli e kelasa kashtavenisitu.	Shama is managing her work very well this week. But once she felt she couldn't manage it. She is looking after her sister's children once in a week without any difficulty. But she felt this task difficult this week.
F	X is a woman in control. She is very self-sufficient and is generally able to handle anything. She prides herself on her ability to handle pressure and hates asking for help from anyone.	Husma thanna hidithadalli iruva mahile. Avalu thumba swaavalambithalu. Dinadalli yene aadaru adannu nirvahisalu saamarthya hondiddale. Avalu otthaddavannu nirvahisuva thanna saamarthyada mele swathaha hemme paduthale mathu bereyavara bali sahaaya kelalu ishta paduvudilla	Shweta is a woman in control. She is very self-sufficient and has is capable of handling anything. She feels proud about her ability to handle stress/pressure and hates asking for help from anyone.

Item 7: I have been so unhappy that I have had difficulty sleeping…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X has been feeling very low recently. This is especially true at night when all the busyness of life stops and her mind is left to wander. She gets stuck in negative thought patterns and lies in bed awake for several hours every night before eventually dropping off.	Varalakshmi avaru itthichige thumba besaara dinda iddare. Dhinada janjaatagalu mugidu kelasa kaarya illadhiro raathriya samaya dalli, ee anubhava hechhuthade. Avaru nakaraathmaka alochaneyelli iruthare, haagu avaru malagidaaga prathi rathri sumaaru gantegala kaala malaguva munna yecharra viruthare.	Varalakshmi is very sad/ upset of late. This feeling increases after finishing daily hassles in the night when there is no work. She will have negative thoughts and every night while going to sleep she will be awake for quite a few hours before she can go to sleep.
B	X has been so unhappy recently that she has only been able to manage a few hours' sleep a night. She wakes up 3 or 4 times a night with great feelings of sadness and disappointment each night. It takes around an hour fall back to sleep again.	Nuthan avaru itthichege yeshtu dhukhadhalli iddhaarendhare avaru kelavu gantegalu maathra raathri niddhe maadalu saadhyavagutthidhe. Raathri moornalku baari niraashe haagu bejaru anubhavisuttha yeccharavagutthaare. Avarige matthe nidhre baralu sumaru ondhu ganteyaaguthadhe.	Nowadays Nuthan is so much so sad that she could sleep only for few hours at night this week. She wakes up out of disappointment and sadness 3 -4 times at night. It will take an hour to catch sleep again.
C	Half of the time, X is able fall asleep easily and sleeps right through the night, but on 3 or 4 nights of the week she is feeling so upset about her life that she cannot sleep easily. On these nights it takes her about an hour to fall asleep.	Ardhadhastu samaya Kavitha raathri idee aaramavaagi nidhre maadalu saadhyavaaguthadhe. Aadhare vaaradalli moorarindha naalku raathrigalu besara dalli mulugiddhu, nidhre maadalu saadhyavaaguvudilla. Inthaha samayagalalli raathri nidre madalu sumaru ondhu gantegala kaala bekaagutthade.	Half of the time, Meena sleeps right through the night, but on 3 or 4 nights of the week she is feeling so upset that she cannot sleep. On such nights it takes her about an hour to fall asleep.
D	X is very unhappy in her job. She only works two days a week, but the thought of going in the next day makes her sleep fitful and unsettled so she does not feel well rested when she wakes up.	Sunandalige avala kelasadha mele thumba asamaadhaanavidhe. Vaaradalli kevala eradu dina maathra kelasakke hogutthaare. Aadhare niddhre maaduvaaga naale dhinadha kelasadha bagge yochane maadutthaa, avaru nidhreyalli chadapadisutthaare. Idharindha beligge yeccharavaadhaaga nidhre maadidha nanthara siguva vishraanthi siguthilla.	Radha is very dissatisfied with her work. She only works two days a week, but the thought of going to work the next day makes her restless during sleep the night before. She therefore does not feel rested and fresh when she wakes up.
E	X generally sleeps very well but there is always one night every week when she struggles to fall asleep. On these nights she feels very depressed and can't stop thinking about all the things that have gone wrong in her life.	Shylaja samaanyavaagi thumba chennagi nidde maaduthale. Aadare vaaradalli ondu rathri nidde madalu kashta paduthale. Aa rathrigalalli avalu tumba khinnathe ge olagadhanthe bhaavisuthale matthu thanna jeevanadalli thappagi nadediruva vishayagala bagge yochisuvadannu avalige nillisalu aaguvadilla.	Sheela usually sleeps well at night but once a week she struggles to sleep. On those nights she feels very depressed and tends to worry about things that have gone wrong in her life and is unable to stop herself from thinking so.
F	X falls asleep within 10 minutes every night. She goes into a deep sleep and does not wake up until her alarm goes off in the morning, feeling well rested and refreshed.	Kavya dina raathri 10 nimishadalli nidde hogutthale. Avalu gaada nidhre maduthale mathu avalu thanna alarm hodeyuva varegoo yeddeluvudilla. Dhanivillade mathu aaramavaagi yeddeluthale.	Shweta falls asleep within 10 minutes every night. She sleeps deep and does not wake up until her alarm goes off. She wakes up, feeling no fatigue and refreshed.

Item 8: I have felt sad or miserable…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X is so unhappy that she tries to avoid people whenever possible and needs to go off to be by herself 3 or 4 times a day. No one and nothing is able to cheer her up.	Hemavathi avaru yeshtu dhukhadallidhaare endhare janarindha aadhashtu dhooraviralu prayathnisutthaare. Avarige dinadalli moornaalku baari ontiyaagi irabekendhu anisutthadhe. Yaarigoo avarannu khushi padisalu saadhyavaaguthilla.	Lakshmi is so sad that she tries to stay away from people whenever possible. She feels like being by herself 3 or 4 times a day. No one is able to cheer her up.
B	X is so unhappy that she finds it difficult to get out of bed every morning. Once she is out of bed though, she is able to distract herself with jobs and people and feels better for the rest of the day.	Almelu avaru eshtu dhukadinda iddare endare dina beligge hasige inda elalu oddaduthare. Aadare yedda nanthara avaru kelasa kaaryadalli thodagi, janarondige bereyuthare. Ulidha samayadalli aaramavaagi iruthare.	Almelu is so sad that she finds it difficult to get up from the bed every morning. But she involves in work and mingles with people once she gets up. She'll be comfortable rest of the time.
C	X enjoys spending time with her family and feels positive about the future. She gets depressed and feels very low 3 or 4 times a week. This causes her to lose interest in the things she normally enjoys.	Nageena thanna kutumbhadavarodane samaya kaleyalu anandhisutthaare haagu thanna bhavishyadha bagge aashavadhiyaagidhaare. Vaaradhalli moorarindha naalku baari avaru khinnathege olagaagi bahala besaradindha iruthaare. Idharinda avaru munche khushi indha maaduthidha kelasa kaaryagalalli aasakthi kaledhu kollutthaare.	Meena enjoys spending time with her family and enjoys. She is also hopeful about her future. She feels very sad and depressed 3 or 4 times a week. This has made her to lose interest in things she would enjoy doing earlier.
D	X is normally a cheerful and outgoing person but twice a week she goes through a time of feeling very unhappy. When this happens she becomes very withdrawn and unsociable. She is able to come out of it after about an hour when she focuses on something enjoyable.	Shabana saamanyavaagi uthsaaha dhindha irutthaare. Aadhare vaaradalli eradu dina athrupthikara bhaavane untaaguthadhe. Ee reethi untaadhaaga avaru obbare irutthaare matthu yaarodaneyu bereyuvudilla. Yaavaga avaru ishtapaduva vichaaradha bagge gamana harisutthaalo, aaga ee reethi varthane indha sumaaru ondhu ganteya nanthara aache baralu saadhyavaaguthadhe.	Radha is usually known to be a happy and enthusiastic person. Couple of times a week she feels very unhappy and dissatisfied. During times like these, she prefers to be alone. She is able to overcome this and focus on something enjoyable only after about an hour
E	X is a happy person who loves spending time with her family and gets to hang out with her friends 2 or 3 times a week. Once a week she feels very low for half an hour but she doesn't see this as particularly remarkable.	Asha santhoshada vyakthi. Thanna kutumbadondige samaya kaleyalu preethisuva vyakthi. Matthu vaaradalli 2 athava 3 baari snehithara jothe samaya kaleyuthale. Vaarakomme ardha ganteyavarege thumba dhuka anubhavisuthale. Aadare avalu idannu nirdishta vaagi gamanisuvudilla.	Sheela is a happy person and likes to spend time with her family. She spends time with friends twice or thrice a week. However once a week for about half an hour she feels very sad but doesn't notice/view this as being significant.
F	X loves life and is always happy. She finds great joy in spending time with her children and finds her work very rewarding.	Gayathri jeevanavannu preethisuthale mathu yaavagalu khushiyagiruthale. Avalu makkala jothe samaya kaleyuttha santoshavannu kandu kollutthale. Avalu thanna kelasadalli puraskaarada anubhava padiyuthalle.	Shweta loves her life and is always happy. She enjoys spending time with her children. She also finds her work very rewarding.

Item 9: I have been so unhappy that I have been crying…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X feels so miserable that she tries to avoid people whenever possible. She breaks down in tears 3 or 4 times a day.	Misba avaru yeshtu sankatadalli iddhaarendhare, avaru janarindha aadhastu dhooraviralu prayathnisuthaare. Dhinadalli moornaalku baari dhuka tadeyalaagadhe aluttha irutthaare.	Lakshmi feels so miserable that she tries to stay away from people whenever possible. She feels so sad that she cries 3 or 4 times a day.
B	X is finding life extremely difficult at the moment. She manages to get through the day by keeping herself busy but breaks down at night when she has nothing to distract her and cries for about an hour before she falls asleep	Asma avirige avara jeevana bahala kashtadinda thumbiruvanthe anisuthade . Avaru kaaryaniratharagi, dina nithyada kelasagallanu nibhaayisikondu hoguthare. Aadare prathi raathri avarige kelasa illadiro saamayadalli malaguva munche sumaaru ondu gante aluthare.	Asma feels her life is full of difficulties/ misery. She proactively and responsibly handles daily chores. But she cries for one hour every night when she didn't have work before going to the bed.
C	Things have been getting X really down recently. She feels like she is growing apart from her husband and is finding her work extremely stressful. This is becoming too much to handle and brings her to tears 3 or 4 times a week.	Shakuntala ge ittichege paristhithigalu hathaashe golisutha idhe. Thanna ganda nindha dhoora aaguthiruvanthe anisuthiddhu, kelasa kaaryagalu otthada endhu anisuttide. Idhannu nibhayisalu kashta vaagiddhu, vaaradalli moorarindha naalku baari aluthaare.	Meena has been feeling really overwhelmed by situations recently. She feels a growing distance with her husband and is finding her work extremely stressful. She finds it hard to manage this and cries 3 or 4 times a week.
D	X enjoys life the most when she is busy. She has no problem with feeling low or sad when she is going about her daily activities or doing one of her hobbies. When all of that stops though she feels very unhappy, and is even brought to tears twice a week.	Shruthi thaanu kelasa kaaryagalli thodagiddhaga haagu chatuvatike indha iruvaaga jeevanadha bagge santhosha paduthaale. Avalu dhina nithyadha chatuvatikeyalli bhaagiyaagiruvaaga yene yeruperaadharu avalu hecchu thale kedisikolluvudilla. Aadhare yaavaga ee yella kaaryagalu nilluvudho aaga avalige hecchu dhukha aaguthadhe. Vaaradalli eradu baari idharinda atthiddaale.	Radha enjoys life the most when she is active and busy. When she is involved in her daily activities, she hardly gets affected by ups and downs. When all of these activities end, she feels very sad, and has cried twice in a week.
E	X is a cheerful individual who loves life. She is outgoing and really enjoys spending time with people. She does feel unhappy at times though, and usually cries once a week.	Swathi jeevanavannu preethisuva santhoshada vyakthi. Avalu sneha jeevi matthu janara jothe samaya kaleyuvudannu anandisuthale. Aadare kelavomme besara anubhavisuthale mathu saamaanyavagi vaarakomme aluthale.	She loves life and is a happy person. She likes spending time with people. She however feels sad at times and has cried once a week.
F	X feels very happy and contented with her life. She sometimes feels a bit low or unhappy but it is not enough to cause her to cry.	Asma thumba khushi paduthale matthu thanna jeevanadalli santhrupthalagiddale. Avalu kelavu baari swalpa besara athava athrupthi hondiddale. Aadare adu avalu alalu karanavagalilla.	Shweta feels very happy and satisfied with her life. She sometimes feels a little sad or dissatisfied but it is not a reason for her to cry.

Item 10: The thought of harming myself has occurred to me…			
AV	**English Final Version**	**Kannada Final Version**	**English Back-Translation**
A	X is currently under an overwhelming level of emotional distress. She worries a lot about money and has recently found out that she is pregnant. She is very anxious about the future and thinks about harming herself 3 or 4 times a day.	Sadhyadha paristhithiyalli Savithri avaru thadeyalaagadhashtu maanasika otthadadalli idhaare. Avaru hanakaasina bagge athiyaagi chinthisutthaare. Itthichege aake gharbhiniyaagiddhaare yendhu thilididhe. Bhavishyadha bagge thumba aathankadalli iddhaare matthe dhinadalli moornaalku baari avarannu haani maadikolluva bagge yochane maadutthaare.	Currently Savithri is under the stress which can't be tolerable. She worries a lot about finance. She recently came to know that she's pregnant. She is apprehensive about her future and she thinks about self harm 3 -4 times in a day.
B	X's husband died a few months ago and ever since she has been struggling to cope with her grief and the pressure of making ends meet by herself. She finds it difficult to talk about how she is feeling and thinks about harming herself once a day.	Kala avara ganda kelavu thingalugala kelage theerikondaru. Andininda e dhukkavannu nibhaayisalu thumba sankatavenisuthade. Haagu maneya javaabdaarigalannu avarobbare hora bekemba otthadalli iddare. Avarige ee vishayada bagge yaara baliyu mathanaadalu kashtavenisuthide, haagu dinadalli ondu baari avarige haani maadikolluva yochane baaruthade.	Kala's husband passed away few months ago. She is finding it difficult to manage her grief since then. And she is under the pressure of taking the family responsibilities all alone. She finds it difficult to discuss this with anyone, and she thinks about self harm once in a day.
C	X is currently very ill and unable to do any of the things she normally enjoys. She also feels completely useless as she cannot help out with the children or go to work. There are 3 or 4 times a week where the thought of harming herself occurs to her.	Sangeethalige itthichege husharilla. Avaru munche khushi indha maadutthidha kelaskaryagalannu eega maadalu saadhyavaaguthilla. Avaru makkalige sahaaya maadalu athava kelasakke hogalu aguthilla. Adharindhaagi thaanu nishprayojaka matthu, vaaradalli moorarinda naalku baari thanage thaanu haani maadikolluva aalochane barutthade.	Sangeetha is not well nowadays. She can't do the work now which she enjoyed to do before. She is unable to help the children or to go for work. Hence she feels that she is useless and she thinks about self harm 3 -4 times in a week.
D	X doesn't feel like she is in control of her life. Her life is so hectic that she feels like she is constantly reacting to one stressful situation after another. It crosses her mind twice a week to self-harm.	Mousina la jeevana thanna hidithadhalli illa yendhu anisutthidhe. Avarige jeevanadalli ondhara melondhu thumba otthadadha sanniveshagalannu anubhavisutthiruvanthe anisuthidhe. Vaaradalli eradu baari aakeya manassige thannanu thanu haani maadikolluva aalochane barutthade.	Mausina feels her life is out of her control. She feels she is facing stressful situations one after another in her life. She thinks about self harm 2 times in a week.
E	X is being bullied by her family. They accuse her of being lazy and not pulling her weight around the house. She feels like she is under constant attack and it has gotten so bad in the last week that at one point, she even thought about harming herself.	Mamathallannu avala kutumbadavaru heeyalisuthare. Avarella avallannu somberi yendu aaropisuthare mathu maneyalli yellaranthe kelasa maduvudilla vendu heluthare. Avarella niranthara bayyutthare yendu avalige anisuthade mathu idu kaleda vaara yeshtu kettadagide yendare ondu hanthadalli avalu thannannu haani madi kolluvudara bagge yochisiddalu.	Sheela's family members criticize/harass her and consider her to be lazy and accuse her of not doing any work in the house like her other family members. She feels that they constantly criticize her and last week it went to an extent where she even thought of harming herself.
F	X is under a lot of pressure at work and it has been very difficult to cope with. Luckily, she has been able to let off steam with her family in the evenings and has not got to the point where she has thought about harming herself.	Hajeera kelasadalli thumba otthaddavide mathu adannu nibaayisalu thumba kashta vaagide. Adrushttavashaath sanje vele avalu thanna bhavaane gallannu kutumbhadondige hanchikolluthale mathu thannannu haani madi kolluva yochane ge innu bandilla.	Shweta is under a lot of pressure at work and it has been very difficult for her to manage it. Fortunately, in the evenings she has been able to share her feelings with her family and has not gone to the point where she has thought of harming herself.

Appendix 9: Measurement Invariance Script

BCHADS Longitudinal Invariance Testing Script

Configural Model

```
gsem (EPDS_T5-> ph5_EPDS_1 ph5_EPDS_2 ph5_EPDS_3 ph5_EPDS_4 ph5_EPDS_5 ph5_EPDS_6 ///
ph5_EPDS_7 ph5_EPDS_8 ph5_EPDS_9 ph5_EPDS_10, family(bernoulli) link(logit)) ///
(EPDS_T8-> ph8_EPDS_1 ph8_EPDS_2 ph8_EPDS_3 ph8_EPDS_4 ph8_EPDS_5 ph8_EPDS_6 ph8_EPDS_7 ///
ph8_EPDS_8 ph8_EPDS_9 ph8_EPDS_10, family(bernoulli) link(logit)), cov(EPDS_T5*EPDS_T8) ///

estimate store configural
```

Metric Model

```
gsem (EPDS_T5-> ph5_EPDS_1@i1 ph5_EPDS_2@i2 ph5_EPDS_3@i3 ph5_EPDS_4@i4 ///
ph5_EPDS_5@i5 ph5_EPDS_6@i6 ph5_EPDS_7@i7 ph5_EPDS_8@i8 ph5_EPDS_9@i9 ///
ph5_EPDS_10@i10, family(bernoulli) link(logit)) ///  (EPDS_T8-> ph8_EPDS_1@i1 ph8_EPDS_2@i2 ///
ph8_EPDS_3@i3 ph8_EPDS_4@i4 ph8_EPDS_5@i5 ph8_EPDS_6@i6 ph8_EPDS_7@i7 ///
ph8_EPDS_8@i8 ph8_EPDS_9@i9 ph8_EPDS_10@i10, family(bernoulli) link(logit)), cov(EPDS_T5*EPDS_T8)

estimate store metric
lrtest configural
```

Partial Metric Model (factor loadings freed for item 2)

```
gsem (EPDS_T5-> ph5_EPDS_1@i1 ph5_EPDS_2@i2 ph5_EPDS_3@i3 ph5_EPDS_4@i4 ph5_EPDS_5@i5 ///
ph5_EPDS_6@i6 ph5_EPDS_7@i7 ph5_EPDS_8@i8 ph5_EPDS_9@i9 ph5_EPDS_10@i10, ///
family(bernoulli) link(logit))  (EPDS_T8-> ph8_EPDS_1@i1 ph8_EPDS_2@m2 ph8_EPDS_3@i3 ///
ph8_EPDS_4@i4 ph8_EPDS_5@i5 ph8_EPDS_6@i6 ph8_EPDS_7@i7 ///
ph8_EPDS_8@i8 ph8_EPDS_9@i9 ph8_EPDS_10@i10, family(bernoulli) link(logit)), cov(EPDS_T5*EPDS_T8)

estimate store partial_metric
lrtest configural
```

Partial Scalar Model (thresholds freed for item 2)

```
gsem (EPDS_T5-> ph5_EPDS_1@i1 ph5_EPDS_2@i2 ph5_EPDS_3@i3 ph5_EPDS_4@i4 ///
ph5_EPDS_5@i5 ph5_EPDS_6@i6 ph5_EPDS_7@i7 ph5_EPDS_8@i8 ph5_EPDS_9@i9 ///
ph5_EPDS_10@i10, family(bernoulli) link(logit))  (EPDS_T8-> ph8_EPDS_1@i1 ph8_EPDS_2@m2 ///
ph8_EPDS_3@i3 ph8_EPDS_4@i4 ph8_EPDS_5@i5 ph8_EPDS_6@i6 ph8_EPDS_7@i7 ///
ph8_EPDS_8@i8 ph8_EPDS_9@i9 ph8_EPDS_10@i10, family(bernoulli) link(logit)) ///
(ph5_EPDS_1<-_cons@c1) (ph5_EPDS_2<-_cons@c2) (ph5_EPDS_3<-_cons@c3) (ph5_EPDS_4<-_cons@c4) ///
(ph5_EPDS_5<-_cons@c5) (ph5_EPDS_6<-_cons@c6) (ph5_EPDS_7<-_cons@c7) (ph5_EPDS_8<-_cons@c8) ///
 (ph5_EPDS_9<-_cons@c9) (ph5_EPDS_10<-_cons@c10) (ph8_EPDS_1<-_cons@c1) (ph8_EPDS_2<-_cons@cc2)
///
 (ph8_EPDS_3<-_cons@c3) (ph8_EPDS_4<-_cons@c4) (ph8_EPDS_5<-_cons@c5) (ph8_EPDS_6<-_cons@c6) ///
(ph8_EPDS_7<-_cons@c7) (ph8_EPDS_8<-_cons@c8) (ph8_EPDS_9<-_cons@c9) (ph8_EPDS_10<-_cons@c10),
/// cov(EPDS_T5*EPDS_T8)

estimate store partial_scalar
lrtest partial_metric
```

**All thresholds significantly different.*

WCHADS Longitudinal Invariance Testing Script

Configural Model.

```
gsem (EPDS_T5-> ph5_EPDS_1 ph5_EPDS_2 ph5_EPDS_3 ph5_EPDS_4 ph5_EPDS_5 ph5_EPDS_6 ///
```

```
ph5_EPDS_7 ph5_EPDS_8 ph5_EPDS_9 ph5_EPDS_10, family(bernoulli) link(logit)) ///
(EPDS_T8-> ph8_EPDS_1 ph8_EPDS_2 ph8_EPDS_3 ph8_EPDS_4 ph8_EPDS_5 ph8_EPDS_6 ///
ph8_EPDS_7 ph8_EPDS_8 ph8_EPDS_9 ph8_EPDS_10, family(bernoulli) link(logit)), cov(EPDS_T5* EPDS_T8)

estimate store configural
```

Metric Model

```
gsem (EPDS_T5-> ph5_EPDS_1@i1 ph5_EPDS_2@i2 ph5_EPDS_3@i3 ph5_EPDS_4@i4 ph5_EPDS_5@i5 ///
ph5_EPDS_6@i6 ph5_EPDS_7@i7 ph5_EPDS_8@i8 ph5_EPDS_9@i9 ph5_EPDS_10@i10, ///
family(bernoulli) link(logit)) (EPDS_T8-> ph8_EPDS_1@i1 ph8_EPDS_2@i2 ph8_EPDS_3@i3 ///
ph8_EPDS_4@i4 ph8_EPDS_5@i5 ph8_EPDS_6@i6 ph8_EPDS_7@i7 ph8_EPDS_8@i8 ///
ph8_EPDS_9@i9 ph8_EPDS_10@i10, family(bernoulli) link(logit)), cov(EPDS_T5*EPDS_T8)

estimate store metric
lrtest configural
```

Scalar Model

```
gsem (EPDS_T5-> ph5_EPDS_1@i1 ph5_EPDS_2@i2 ph5_EPDS_3@i3 ph5_EPDS_4@i4 ///
ph5_EPDS_5@i5 ph5_EPDS_6@i6 ph5_EPDS_7@i7 ph5_EPDS_8@i8 ph5_EPDS_9@i9 ///
ph5_EPDS_10@i10, family(bernoulli) link(logit)) (EPDS_T8-> ph8_EPDS_1@i1 ph8_EPDS_2@i2 ///
ph8_EPDS_3@i3 ph8_EPDS_4@i4 ph8_EPDS_5@i5 ph8_EPDS_6@i6 ph8_EPDS_7@i7 ///
ph8_EPDS_8@i8 ph8_EPDS_9@i9 ph8_EPDS_10@i10, family(bernoulli) link(logit)) ///
(ph5_EPDS_1<-_cons@c1) (ph5_EPDS_2<-_cons@c2) (ph5_EPDS_3<-_cons@c3) (ph5_EPDS_4<-_cons@c4) ///
(ph5_EPDS_5<-_cons@c5) (ph5_EPDS_6<-_cons@c6) (ph5_EPDS_7<-_cons@c7) (ph5_EPDS_8<-_cons@c8) ///
(ph5_EPDS_9<-_cons@c9) (ph5_EPDS_10<-_cons@c10) (ph8_EPDS_1<-_cons@c1) (ph8_EPDS_2<-_cons@c2)
///
(ph8_EPDS_3<-_cons@c3) (ph8_EPDS_4<-_cons@c4) (ph8_EPDS_5<-_cons@c5) (ph8_EPDS_6<-_cons@c6) ///
(ph8_EPDS_7<-_cons@c7) (ph8_EPDS_8<-_cons@c8) (ph8_EPDS_9<-_cons@c9) (ph8_EPDS_10<-_cons@c10),
/// cov(EPDS_T5*EPDS_T8)

estimate store scalar
lrtest metric
```

Partial Scalar Model (item 5, 9, 7, 4 freed)

```
gsem (EPDS_T5-> ph5_EPDS_1@i1 ph5_EPDS_2@i2 ph5_EPDS_3@i3 ph5_EPDS_4@i4 ph5_EPDS_5@i5 ///
ph5_EPDS_6@i6 ph5_EPDS_7@i7 ph5_EPDS_8@i8 ph5_EPDS_9@i9 ph5_EPDS_10@i10, ///
family(bernoulli) link(logit)) (EPDS_T8-> ph8_EPDS_1@i1 ph8_EPDS_2@i2 ph8_EPDS_3@i3 ///
ph8_EPDS_4@i4 ph8_EPDS_5@i5 ph8_EPDS_6@i6 ph8_EPDS_7@i7 ///
ph8_EPDS_8@i8 ph8_EPDS_9@i9 ph8_EPDS_10@i10, family(bernoulli) link(logit)) ///
(ph5_EPDS_1<-_cons@c1) (ph5_EPDS_2<-_cons@c2) (ph5_EPDS_3<-_cons@c3) (ph5_EPDS_4<-_cons@c4) ///
(ph5_EPDS_5<-_cons@c5) (ph5_EPDS_6<-_cons@c6) (ph5_EPDS_7<-_cons@c7) (ph5_EPDS_8<-_cons@c8) ///
(ph5_EPDS_9<-_cons@c9) (ph5_EPDS_10<-_cons@c10) (ph8_EPDS_1<-_cons@c1) (ph8_EPDS_2<-_cons@c2)
///
(ph8_EPDS_3<-_cons@c3) (ph8_EPDS_4<-_cons@cc4) (ph8_EPDS_5<-_cons@cc5) (ph8_EPDS_6<-_cons@c6)
///
(ph8_EPDS_7<-_cons@cc7) (ph8_EPDS_8<-_cons@c8) (ph8_EPDS_9<-_cons@cc9) (ph8_EPDS_10<-
_cons@c10), /// cov(EPDS_T5*EPDS_T8)

estimate store partial_scalar
lrtest metric
```

T5 Multigroup Invariance Testing Script

** Configural Model*

```
gsem (2: L1 -> ph5_EPDS_1@b1 ph5_EPDS_2@b2 ph5_EPDS_3@b3 ph5_EPDS_4@b4 ph5_EPDS_5@b5 ph5_EP
///
DS_6@b6 ph5_EPDS_7@b7 ph5_EPDS_8@b8 ph5_EPDS_9@b9 ph5_EPDS_10@b10, family(ordinal) link(logit))
///
(1: L1 -> ph5_EPDS_1@b11 ph5_EPDS_2@b12 ph5_EPDS_3@b13 ph5_EPDS_4@b14 ph5_EPDS_5@b15 ///
ph5_EPDS_6@b16 ph5_EPDS_7@b17 ph5_EPDS_8@b18 ph5_EPDS_9@b19 ph5_EPDS_10@b20,
family(ordinal) /// link(logit)), variance(L1@1) mean(L1@0) group(cohort) ginvariant(none)byparm latent(L1)
nocapslatent

estimate store configural
```

**Partial Metric model (3, 4, 7, 8, 9, 10 freed)*

```
gsem (2: L1 -> ph5_EPDS_1@b1 ph5_EPDS_2@b2 ph5_EPDS_3@b3 ph5_EPDS_4@b4 ph5_EPDS_5@b5 ///
ph5_EPDS_6@b6 ph5_EPDS_7@b7 ph5_EPDS_8@b8 ph5_EPDS_9@b9 ph5_EPDS_10@b10, family(ordinal) ///
 link(logit)) (1: L1 -> ph5_EPDS_1@b1 ph5_EPDS_2@b2 ph5_EPDS_3@b13 ph5_EPDS_4@b14 ph5_EPDS_5@b5
///
ph5_EPDS_6@b16 ph5_EPDS_7@b17 ph5_EPDS_8@b18 ph5_EPDS_9@19 ph5_EPDS_10@b11, family(ordinal)
/// link(logit)), variance(L1@1) mean(L1@0) group(cohort) ginvariant(none)byparm latent(L1) nocapslatent

estimate store partial_metric
lrtest configural
```

**Minimum requirements for partial metric not met so did not proceed to test scalar model.*

T8 Multigroup Invariance Testing Script

**Configural Model*

```
gsem (2: L1 -> ph5_EPDS_1@b1 ph5_EPDS_2@b2 ph5_EPDS_3@b3 ph5_EPDS_4@b4 ph5_EPDS_5@b5 ///
ph5_EPDS_6@b6 ph5_EPDS_7@b7 ph5_EPDS_8@b8 ph5_EPDS_9@b9 ph5_EPDS_10@b10, ///
family(ordinal) link(logit)) (1: L1 -> ph5_EPDS_1@b11 ph5_EPDS_2@b12 ph5_EPDS_3@b13 ph5_EPDS_4@b14
/// ph5_EPDS_5@b15  ph5_EPDS_6@b16 ph5_EPDS_7@b17 ph5_EPDS_8@b18 ph5_EPDS_9@19 ///
ph5_EPDS_10@20, family(ordinal) link(logit)),   variance(L1@1) mean(L1@0) group(cohort) ///
ginvariant(none)byparm latent(L1) nocapslatent

 estimate store configural
```

**Partial metric model (1, 2, 6 & 8 freed)*

```
gsem (2: L1 -> ph8_EPDS_1@b111 ph8_EPDS_2@b222 ph8_EPDS_3@b3 ph8_EPDS_4@b4 ph8_EPDS_5@b5 ///
ph8_EPDS_6@b666 /// ph8_EPDS_7@b7 ph8_EPDS_8@b888 ph8_EPDS_9@b9 ph8_EPDS_10@b10,
family(ordinal) /// link(logit)) /// (1: L1 -> ph8_EPDS_1@b11 ph8_EPDS_2@b2 ph8_EPDS_3@b3
ph8_EPDS_4@b4 ph8_EPDS_5@b5 /// ph8_EPDS_6@b16 ph8_EPDS_7@b7 ph8_EPDS_8@b18
ph8_EPDS_9@b9 ph8_EPDS_10@b10, family(ordinal) ///
link(logit)),  variance(L1@1) mean(L1@0) group(cohort) ginvariant(none)byparm latent(L1) nocapslatent

18 . estimate store partial_metric
19 . lrtest configural
```

*Test for unequal thresholds

```
. forvalues i=1/10{
qui: test _b[/ph8_EPDS_`i':1.cohort#c.cut1]=_b[/ph8_EPDS_`i':2.cohort#c.cut1]
di "Factor 1 - Item " `i' " cut 1: chi2=" r(chi2) , "pvalue =
" r(p)
}
```

*8/10 thresholds significantly different so no further testing.