Working Memory Deficits in Chronic Pain

Christopher Knaggs

Supervised by:

Dr Christopher Brown

Dr Hannah Twiddy

Dr Monika Halicka

June 2021

Submitted in partial fulfilment of the Doctorate in Clinical Psychology, University of Liverpool

Acknowledgements

There are many people that I am grateful to for supporting me to conduct this thesis. My supervisors Chris Brown, Monika Halicka and Hannah Twiddy have been a constant source of knowledge, guidance and encouragement. They have always responsive and keen to help, which has been vital as I have navigated the impact of COVID-19 on my project alongside increased childcare and compacted time for other aspects of the course. I am so grateful for their flexibility and willingness to find time to push their support for this project up their priority list in order to complete it in time.

I am grateful to everybody who took part in the empirical study. Each participant completed the experiment in their own home and in their own time, requiring personal motivation and dedication to contributing to research which helps those living with CP, regardless of whether this affected their life or not personally. I was struck by the high proportion of people who simply wanted to help those in pain. I am grateful to my friends and family, for both emotional and practical support. Completing this project during COVID-19 has been stressful and I am grateful as always to my wife and son for their love, patience and humour, without which it would have been much harder. I have not seen much of my other family or friends due to restrictions, but when possible others have provided childcare when I needed to find time to work, taken part in pilot experiments, spread my recruitment needs via word of mouth, or been there to talk when I needed them. I would be nowhere near completing this study on time, and would be in a much worse place psychologically without any of them. Similarly, although I have sadly seen very little of my cohort throughout most of this project, we have been in touch by phone and text, and I am grateful for their advice, validation and reassurance.

Table of Contents

List of Tables

Chapter 1: Literature Review

Chapter 2: Empirical Paper

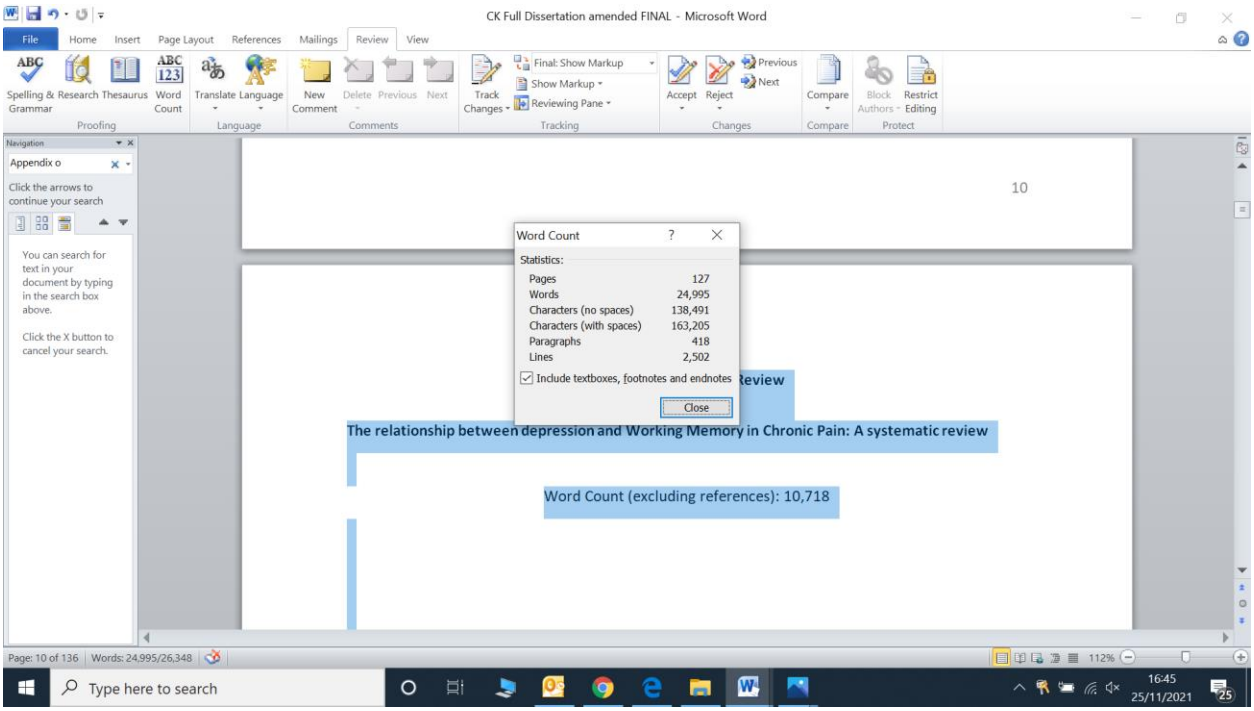
List of Figures

Chapter 1: Literature Review

Chapter 2: Empirical Paper

List of Appendices

Word Count: 24,995



Introductory Chapter: Thesis Overview

This thesis explores the role of Working Memory (WM), a cognitive process, in contributing to Chronic Pain (CP) conditions, and the emotional and functional difficulties experienced by people with a CP diagnosis. This thesis is formed of two separate but related papers: (1) a systematic review examining the current evidence in the literature for a relationship between WM impairment and depression in people with a diagnosis of a CP condition; and (2) an original quantitative study which seeks to validate a novel measure of WM, while also providing pilot data concerning potential differences in WM functioning between participants with Fibromyalgia (FM), a CP condition, and Healthy Controls (HC).

Chapter 1: The systematic literature review explores the relationship between WM impairments and depression in people with a diagnosis of a CP condition. Four databases (Medline, Pubmed, PsychInfo and CINAHL) were searched, concluding in February 2021. Ten studies were identified and underwent quality assessment of methodology. Nine out of the ten studies did not report a statistically significant relationship between WM and depression outcomes, and one did. However, there were issues with methodology, reporting of outcomes and sample sizes, which may have led to several relevant outcomes being omitted or unacknowledged. When outcomes other than statistical significance were considered, there was more evidence of a level of association between WM and depression in some subgroups when arranged by type of CP diagnosis and sensory modality of WM measure. The methodological issues and theoretical and clinical implications were discussed with respect to these findings.

Chapter 2: The empirical study aimed to validate a novel measure of a subtype of WM called Auditory WM (AWM). This study also provided pilot data concerning differences in AWM capacity in participants with FM, compared to HCs. All participants undertook a series of computer-based tasks which place certain demands of WM and AWM. Quantitative data was gathered and analysed in the form of task

accuracy and reaction times across and within study tasks. While not all of the hypothesised findings consistently emerged, there were sufficient consistent findings to validate the measure. An overall significant difference between two conditions in a within-subjects paradigm was found as hypothesised when the FM and HC groups were combined, but not in each group individually. A lack of between-groups differences and similar effect sizes and means in each group separately and combined suggest that a larger sample size in each group may have been needed to detect the medium effect size across the individual groups. The implications of these findings for further research are discussed, as well as potential modifications that can be made to detect differences between conditions where no trends in line with outcomes were found. There was further analysis of the relationship between clinical and experimental measures of WM, which was found to be in line with expectations. The relevance of all of the study's findings for clinical practice, as well as study limitations are discussed.

Chapter 1: Literature Review

The relationship between depression and Working Memory in Chronic Pain: A systematic review

Word Count (excluding references): 10,718

Paper prepared for submission to the Frontiers in Pain journal for peer review. Author guidelines are in Appendix A.

Abstract

Background: Depression and Working Memory problems are highly prevalent in Chronic Pain populations. Despite the evidence from other clinical groups that depression is positively associated with cognitive deficits including Working Memory, their relationship in chronic pain remains unclear.

Objective: This systematic review aims to synthesise the available evidence for the relationship between Working Memory and Depression in Chronic Pain. It also explores this relationship with regard to different types of Chronic Pain and Working Memory modalities.

Methods: A literature search was carried out using PubMed, PsycINFO, CINAHL and Medline. Eligible studies were observational, used validated quantitative measures of Working Memory and depression and assessed a statistical association between these measures. Participants needed to be aged between 18-64 on average and meet diagnostic criteria for a Chronic Pain condition, or be under the care of a specialist pain management service.

Results: Initial search returned 289 papers. Of these, 10 were included in the final analysis after screening. Most studies were of moderate-to-high quality according to modified Newcastle-Ottawa criteria. Only one paper reported a statistically significant positive association between depression and Working Memory deficits. Further three papers reported some indication of a positive relationship between depression and Working Memory deficits, while not reporting statistical significance. Sample sizes, study design and methodological quality were considered during the results synthesis. Papers which explored Chronic Primary (as compared to secondary) Pain, and which used a visual (as compared to auditory-verbal) measure of Working Memory, were more likely to produce some evidence of a relationship.

Conclusion: While overall there was no evidence of a relationship between Working Memory and Depression in Chronic Pain, subgroup analyses suggest that there may be a positive relationship

between these factors in Chronic Primary Pain samples, and when visual measures of Working Memory are used. However, these preliminary conclusions are limited due to a small number of studies, insufficient statistical power to detect small or medium effect sizes, and methodological issues. Higher quality research is needed to understand the relationship between different Chronic Pain diagnoses and modalities of Working Memory assessment.

Keywords: Chronic Pain, Working Memory, depression, visual, auditory-verbal

Introduction

Chronic pain (CP) is defined as pain which persists for more than three months and significantly impacts on functionality or emotional well-being (Treede et al., 2019). Functionality according to the same paper is defined as 'interference with activities of daily life and participation in social roles', but does not specifically mention cognitive deficits and whether these might be considered factors which reflect, or mediate, difficulties with day-to-day functioning. Guidelines for CP clinical trials outline outcome domains for consideration, which include only physical and emotional factors (Dworkin et al., 2005). The role of cognition in CP, and how it interacts with emotional well-being is not as well understood or incorporated into CP treatment or research frameworks. However, there is preliminary evidence that cognitive deficits play an important role in CP and are related to emotional well-being in general.

Pervasive, distressing or debilitating negative emotions are widely reported by people living with CP (Varni et al., 1996). Anxiety and depression are particularly prevalent issues, with Sagheer, Khan & Sharif (2013) finding that between 49% and 55% of people with a diagnosis of Chronic Low Back pain experienced elevated levels of anxiety and depression respectively. Furthermore, Benjamin et al. (2000) found that people in the UK with a diagnosis of CP are 3.2 times more likely than the general population to have a mental disorder. The sensation of pain itself can induce strong emotional responses (Veinante, Yalcin & Barrot, 2013), while the impact of CP on daily function, relationships and independence further contributes to ongoing heightened emotional arousal (McCracken et al., 1999). It is understandable therefore that CP would be associated with higher levels of emotional distress, with Sturgeon et al (2015) finding a significant correlation between pain intensity and both depression and anger in a cohort of 675 CP patients. Some researchers have developed models which integrate depression and CP into 'somato-affective syndromes', reflecting the close relationship between these phenomena (Peveler, Katona & Wessely, 2006).

Cognitive deficits are widely self-reported by people with CP (e.g. Schnurr & McDonald, 1995). This has been corroborated to an extent by clinician observations, and supported in some cognitive domains by neuropsychological assessment (Dick & Rashiq, 2007). There is further evidence that CP can lead to structural changes in the brain, such as grey matter atrophy, across multiple regions involved in a range of cognitive functions (Bushnell, Ceko & Low, 2013). While there is no literature regarding the wider psychosocial impact of cognitive deficits in CP and how this affects outcomes in terms of functioning, there is substantial evidence that cognitive deficits can adversely affect functionality and quality of life in other populations living with cognitive impairment, such as those with neurological disease (Mitchell et al., 2010).

Among the cognitive domains often identified as being impaired in CP, working memory (WM) is well-researched, with consistent evidence of reduced capacity in CP populations relative to pain-free individuals (Berryman et al., 2013). WM refers to the ability to hold information in mind for short periods (milliseconds to minutes) and manipulate this information. This ability is central to several other cognitive domains, including executive functioning, attention, language and long-term memory (Baddeley, 2010). It is not clear from the existing literature to what extent WM deficits in CP are related to pain itself, or to what extent other factors such as emotional distress or a predisposition to WM deficits are involved.

While a range of types of emotional distress are negatively related to cognitive function (Hart, Wade & Martelli, 2003), depression has been shown to be associated with particularly debilitating cognitive difficulties. There is evidence of depression being associated with poorer memory, attention, executive functioning and WM than HCs (Rock, Roiser, Riedel & Blackwell, 2013; Rose & Ebmeier, 2006).

Additionally, while other emotions are associated with cognitive difficulties – for instance there is evidence that anxiety is associated with attention deficits (Pacheco-Unguetti, Acosta, Callejas & Lupiáñez, 2010) – the relationship between depression and cognition is particularly profound. A meta-analysis by Rock et al. (2013) found evidence that cognitive impairment persists after remission from depression, with a conclusion that this is a continuation of cognitive deficits which are present during active periods of depression. Depression has also been found to be associated with altered brain structure, for instance atrophy of the hippocampus (Sheline, Wang, Gado, Csernansky & Vannier, 1996). Furthermore, Christopher and MacDonald (2005) found that people with clinical depression performed significantly more poorly on a test of WM than HCs or people with an anxiety disorder. These findings combined suggests that there is a particular pervasiveness and clinical significance to the cognitive difficulties associated with depression, and that WM is vulnerable to these effects.

The experience of cognitive difficulties can be distressing for people with CP (Jamison, Sbrocco & Parris, 1989). Living with lasting, recurrent cognitive difficulties (which are often reported by people with CP) is associated with low mood, as found in other populations such as traumatic brain injury (Silver, McAllister & Arciniegas, 2009). Furthermore, there is evidence that cognitive difficulties during remission from depression increase vulnerability to future depressive episodes (Papazacharias & Nardini, 2012) and that WM deficits may contribute to negative cognitive biases commonly found in Major Depressive Disorder (Gotlib & Joormann, 2010). Overall, the literature indicates a consistent relationship between depression and cognitive difficulties, but the evidence is correlational and the issue of causality remains an open question.

While there is considerable literature concerning WM deficits and depression in CP populations, the direct relationship between these phenomena frequently occurring in CP has not been systematically

reviewed. This is important to understand, because the extant evidence for an association between depression and WM has focused on other populations, whereas CP has a distinct clinical profile, with unique factors contributing to difficulties with depression and WM. For instance, persistent pain itself is known to interfere with cognition including WM (Antepohl et al., 2003) and contribute to low mood (Veinante et al., 2013; McCracken et al., 1999), while hypervigilance linked to pain expectation can contribute to WM impediments (Legrain et al., 2011) and emotional distress (Crombez et al., 2005). As such, there are clinical features specific to CP which are not routinely found in other populations where a relationship between WM and depression has already been demonstrated, and the literature from other populations cannot therefore be confidently generalised. What is not currently known is whether depression and WM difficulties are related in CP, as in other populations, or whether no relationship exists, which would suggest that other factors may be more relevant for understanding WM deficits and depression as separate entities. A deeper understanding of this relationship would elucidate this issue and indicate appropriate targets for research into WM difficulties and depression in CP as a distinct clinical group.

In addition, it is not known whether any relationship between depression and WM deficits affects people with specific CP diagnoses, or CP generally. CP conditions have been recently subdivided into two distinct categories: Chronic Primary Pain (CPP) and Chronic Secondary Pain (CSP; Nicholas et al., 2019). While both types involve pain which persists or recurs for longer than three months, the difference is that CSP conditions originate from a physical trauma or illness, while CPP conditions do not (Treede et al., 2019). An example of CPP is fibromyalgia, in which there is widespread pain in the absence of a clear underlying pathology, while an example of CSP is osteoarthritis, in which pain is secondary to degradation of knee cartilage over time (NICE, 2021; NHS, 2019). This distinction is new and there is no current research on the different cognitive profiles of these distinct types of CP, but it has been

proposed that nociplastic pain may explain the mechanisms underlying CPP, but not CSP (Treed et al., 2019). Nociplastic pain is not fully understood, but is different from pain which is adequately accounted for by an underlying pathology, with current conceptualisations focusing on the role of functions of the central nervous system (CNS) in nociplastic pain, for instance sensory processing and augmentation of pain in the CNS (Fitzcharles et al., 2021). Moreover, nociplastic pain can be accompanied by cognitive and affective symptoms associated with CNS activity, such as fatigue, low mood and memory problems (Fitzcharles et al., 2021). As such, there are different mechanisms which have been proposed to underpin CPP and CSP which can be related to CNS activity including cognitive-affective processes. The current systematic review will seek to contribute to the developing area of research differentiating the profiles of CPP and CSP by examining whether any relationship between WM difficulties and depression is more evident in either type of diagnosis. As the literature in this area is limited, this analysis will be exploratory rather than suggesting a specific hypothesis, although based on the limited literature we might expect a more pronounced positive relationship between WM deficits and depression in CPP than CSP.

The evidence outlined thus far raises the possibility that people with CP are stuck in a vicious cycle of depression and cognitive difficulties. A systematic review by Wu et al. (2018) regarding cognitive deficits in FM included a secondary analysis which indicated that severe emotional distress, not specific to depression, is associated with increased general cognitive impairment in FM patients when compared to HCs. The current systematic review intends to build on these findings and provide the first synthesis of the relevant literature on the relationship between depression and WM specifically in CP.

Aims and Objectives

There is currently a gap in knowledge regarding whether depression and WM are associated in CP, and whether there is any evidence of diagnosis-specificity with regard to this relationship. This review aims to systematically evaluate the relationship between WM and depression in CP, in order to address this gap in knowledge. The primary objective was to test the hypothesis that there is a significant positive association between WM deficits and depression in people with CP conditions, and to test whether this relationship is consistent across sub-groups based on type of CP diagnosis and type of WM measure.

Method

This review was carried out in line with guidance from the Centre for Reviews and Dissemination (2009), in addition to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher, Liberati, Tetzlaff, Altman & the PRISMA group, 2009).

Search Strategy

A systematic literature search was undertaken in January 2021 following consultation with an expert librarian in Mersey Care NHS Foundation Trust. Four electronic databases (Scopus, PsycINFO, Medline and CINAHL) were used to search for published literature from any date. Keywords, titles and abstracts were searched for the following terms: ((chronic OR persist* OR “long-term” OR “long term”) AND pain) AND (depress* OR “low mood” OR “mood disorder” OR “affective disorder”) AND (Memory AND (working OR “short-term” OR “short term”)). Only peer-reviewed journal articles were reviewed to increase the likelihood of high quality papers, and only English language papers were considered as translation resources were not available.

The following inclusion criteria were used: (1) observational studies or any study design with an observational component; (2) studies which recruited a clinical sample of people with a diagnosis of a CP

condition with a mean age between 18-64; (3) studies which reported clearly defined quantitative data on WM (validated neuropsychological assessments, cognitive screening tools, self-report scales such as Likert scales or lab-based computerised measures producing outcomes such as accuracy or reaction times) and depression (self-report and clinician rating measures which are validated for use in mental health services); (4) studies which reported outcomes on associations between WM and depression in participants with CP.

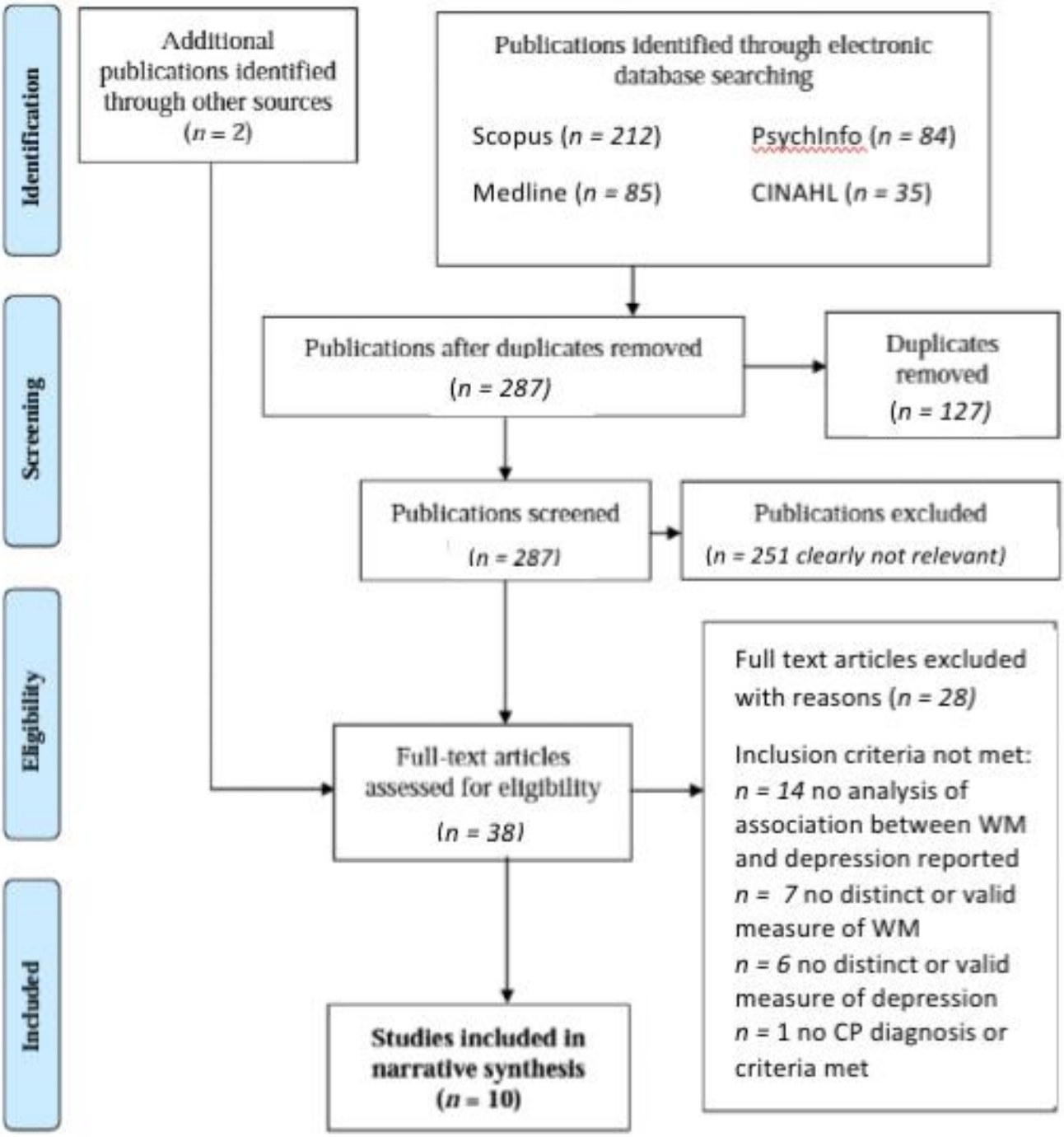
The following exclusion criteria were applied: (1) case reports, thesis/dissertations, book chapters, conference papers, review articles or qualitative papers; (2) studies which did not recruit clinical samples or clinical samples which had (i) no formal diagnosis of a chronic primary pain condition or (ii) not meeting recognised national/international criteria for a CP condition or (iii) not under the care of a pain management service where diagnosis not explicitly stated; (3) studies which recruited clinical participants with a mean age of under 18 or over 64 years of age, to control for developmental or aging effects on cognition; (4) any statistical analysis which did not include a regression or correlation component; (5) studies which used measures of depression which have not been validated for use in clinical settings, or with no normative data.

Study Selection

The first search of the electronic databases identified 414 papers in total. Following removal of duplicates, this number fell to 287. Two authors (First author C.K. and secondary author M.H.) applied the inclusion and exclusion criteria to the titles and abstracts in order to screen for suitability. In line with guidance by Boland, Cherry and Dickson (2017) the secondary author M.H. screened a random sample of 10% of the titles and abstracts (n=29). There were zero disagreements concerning these 29 titles and abstracts, and the screening of the full 287 titles and abstracts by primary author C.K. led to 36

papers to be included in the full text analysis. Full text versions of the 36 selected studies were subsequently acquired and assessed by the two authors independently. Two further studies were identified by manually searching the references lists of the acquired full text papers and screened for eligibility for this study. The study selection protocols outlined above resulted in 10 papers which met the criteria for inclusion in this review (Baker, Gibson, Georgiou-Karastianis & Giummarra, 2018; Dick, Eccleston & Crombez, 2002; Elvemo, Landro, Borchgrevink & Haberg, 2015; Grossi et al., 2008; Herbert et al., 2018; Latysheva, Filativa, Osipova & Danilov, 2019; Oncu, Iliser & Kuran, 2015; Pidal-Miranda, Gonzalez-Villar, Carrillo-de-la-Pena, Andrade & Rodriguez-Salgado, 2018; Radanov et al., 1999; Torkamani et al., 2015). An additional study (Gelonch, Garolera, Valls, Rossello & Pifarre, 2017) was considered potentially eligible if a supplemental data analysis document was made available by the authors, who were contacted regarding this, but no reply was received. An overview of the study selection process is outlined in Figure 1, using the PRISMA flow chart (Moher et al, 2009).

Figure 1. Study selection flow chart (Moher et al., 2009).



Data Extraction and Analysis

A bespoke data extraction table was developed and piloted (see Appendix B) to gather the following information from the final included studies: author; year of publication; type of paper (abstract or article); setting; study design; number of participants; study sponsorship; participant demographics; specific criteria or diagnoses met for CP and clinical characterisations such as time since diagnosis in CP groups, comorbid physical or mental health difficulties reported in study groups; measurement tools, statistical analyses, primary and secondary findings. Effect sizes and statistical significance were extracted if reported, otherwise effect sizes were calculated using available data. This step was conducted by C.K. and cross-checking was completed by supervisor M.H. for accuracy. Minor disagreements in the decisions made were resolved through discussion and reaching consensus.

Assessment of Methodological Quality

Quality assessment was undertaken separately and independently by the two authors already mentioned (C.K. and M.H.). A modified version of the Newcastle-Ottowa Quality Assessment Scale (NOS; Wells, 2014) was used for this purpose (see Appendix C). Discrepancies were resolved through discussion and reaching consensus. This scale uses a star-awarding system from 0-10 stars where zero is lowest quality and 10 is highest quality. It is comprised of four subscales: selection, comparability, exposure, and outcome assessment. It should be noted that a universal scoring system for the NOS has not been established yet (Wang et al., 2018). 0-3, 4-7, and 8-10 stars indicate that a study has low, moderate, or high quality respectively. The assessment of methodological quality produced outcomes where one study attained two stars (Radanov et al., 1999), five studies attained 6-7 stars (Baker et al., 2018; Elvemo et al., 2015; Grossi et al., 2008; Oncu et al., 2015; Torkamani et al., 2015) and four studies were assigned 8-10 stars (Dick et al., 2002; Herbert et al., 2018; Latysheva et al., 2019; Pidal-Miranda et

al., 2018; Appendix D). This suggests that nine out of the 10 studies were of moderate to high quality, while one was low quality.

The clinical group sample sizes of each study may be informative for interpretation of these outcomes during the later Discussion section. A power calculation conducted for this study indicated studies required n=194 to detect a small effect size, n=85 to detect a medium effect size, and n=29 to detect a large effect size through correlational analyses. No studies had a large enough sample size to detect a small effect size and only three were sufficiently powered to detect a medium effect size. Three studies were insufficiently powered to detect even a large effect size (Elvemo et al., 2015 (n=20); Radanov et al., 1999 (n=21); Torkamani et al., 2015 (n=11)).

Results

Study Characteristics

Table 1 outlines the study characteristics of the 10 included studies. Eight of the studies used a cross-sectional design, while two used cohort. Two studies were located in the UK, and one in each of Australia, Canada, the USA, Switzerland, Norway, Russia, Turkey and Spain. Three studies involved a HC group, one study involved a clinical control group, one involved both a HC and a clinical control group, and five studies involved no control group. Cognitive performance was compared against normative data in studies without control groups. Across the ten studies there were 633 participants with CP, 146 HCs and 137 clinical controls recruited. CP sample sizes ranged from 11 to 144. There were a range of CP conditions across the studies, with some studies including multiple conditions. There was a Musculoskeletal Pain sample in three studies, a FM sample in three studies, and a sample of each of the following conditions in one study each: Chronic Back Pain; Whiplash Syndrome; Rheumatoid Arthritis; Idiopathic Pain; Visceral Pain; Irritable Bowel Syndrome; Temporomandibular Pain; Osteoarthritis; Degenerative Disc Disease; Chronic Migraine; Chronic Cluster Headache.

Six of the studies recruited participants who met internationally recognised criteria or diagnostic thresholds for their CP condition either prior to the study or during screening (Dick et al., 2002; Grossi et al., 2008; Latysheva et al., 2019; Oncu et al, 2015; Pidal-Miranda et al., 2018; Torkamani et al., 2015). Two studies indicated that participants with CP were under the care of a pain management or other medical department with expertise in pain management and diagnosis (Elvemo et al., 2015; Radanov et al., 1999). One study stated diagnoses and the need for these diagnoses as inclusion criteria but the criteria used were not made explicit (Baker et al., 2018). In one study CP was verified by a specialist physician based on medical records and assessment (Herbert et al., 2018).

WM was measured using a wide range of clinical and experimental assessments across studies, with some studies using multiple measures. The California Verbal Learning Test was used in two studies (Grossi et al., 2008; Herbert et al., 2018), the N-back task was used in two studies (Elvemo et al., 2015; Radanov et al., 1999), and the following measures of WM were used in one study each: Behaviour Rating Inventory of Executive Function – Adult Version (BRIEF-A) WM subscale (Baker et al., 2018); Test of Everyday Attention (TEA) auditory-verbal WM domain (Dick et al., 2002); Paced Visual Serial Addition Test (PVSAT; Elvemo et al., 2015); Trigram Auditory Memory Task (Grossi et al., 2008); Weschler Adult Intelligence Scale Fourth Edition (WAIS-IV) Letter-Number Sequencing subtest (Herbert et al., 2018); Weschler Adult Intelligence Scale Third Edition (WAIS-II) Spatial Localisation subtest (Pidal-Miranda et al., 2018); Rey Auditory Verbal Learning Test (RAVLT; Latysheva et al., 2019; Oncu et al., 2015); Short Test of Mental Status (STMS; Oncu et al., 2015); WAIS-III WM Index (Torkamani et al., 2015); Warrington Short Recognition Memory for Faces (WSRMF; Torkamani et al., 2015).

There was also some variation in the measures of depression used, with seven studies using the Beck Depression Inventory (Baker et al., 2018; Elvemo et al., 2015; Grossi et al., 2008; Oncu et al., 2015; Pidal-Miranda et al., 2018; Radanov et al., 1999; Torkamani et al., 2015), two studies using the Hospital

Anxiety and Depression Scale depression subscale (Dick et al., 2002; Latysheva et al., 2019) and one study using the Patient Health Questionnaire (Herbert et al., 2018).

The wide range of measures used for depression and particularly WM mean that there is likely to be a limit on effect size accumulation caused by high heterogeneity and low precision (Scammacca, Roberts, & Stuebing, 2014). Consequently a meta-analytical approach was considered unreliable and the results were synthesised narratively.

Table 1. Study Characteristics

Authors (Year)	**Location**	**Study Design**	**Clinical Sample Group Characteristics**	**Clinical Sample Demographics**	**Control Group Characteristics**	**Clinical Sample Diagnostic or Assessment Criteria**
Baker et al. (2017)	Australia	Cross-sectional	CBP including widespread pain n=41: CBP with back pain (25.6%); CBP with other localised pain (53.8%); CBP with widespread pain (15.4%); CBP with joint pain (12.8%); CP duration mean (SD)=12.15 (10.47) years	Mean age (SD)=42.97 (12.77), range 22-65; 24 (61.5%) female; Education mean (SD)=14.74, (3.73) years; 25 (64.1%) tertiary education level; nearly half not working/studying due to pain; Mean estimated IQ (SD): 102.08 (11.21)	n/a	Chronic Back Pain condition necessary for inclusion and pain assessed as part of the study. Exact criteria used not explicitly stated.
Dick et al. (2002)	UK	Cross-sectional, controlled	Chronic persistent pain (FM (n=20), RA (n=20), MSK (n=20)) for at least six months; Mean years' duration (SD): RA=18.9 (15.3); FM=11 (8.6); MSK=10.2 (12.2)	Mean age (SD): RA=62.9 (10.9); FM=48 (16.9); MSK=52.3 (13.1). 16 (80%) female (RA), 18 (90%) female (FM), 12 (60%) female (MSK); Education mean years (SD): RA=12.5 (2.7); FM=13.1 (3.4); MSK=12.7(3.9)	HCs (n=20)	Diagnosis based on medical chart and reassessed, diagnosis made according to American College of Rheumatology criteria
Elvemo et al. (2015)	Norway	Cross-sectional, controlled	Total participants with CP recruited n=20. N=10 (50%) MSK; n=4 (20%) IP; n=1 (5%) VP. Five participants (25%) diagnoses were not reported. Minimum six months with average pain intensity of at least four on the Verbal Rating Scale	Mean age (SD)=38.6 (7.2), range=22 – 49; 16 (80%) of original number recruited female; Mean years education after high school (SD)=4.5 (2.4), range 0 – 10 years	HCs (n=20 initially, 17 after attrition)	Patients under the care of a hospital pain clinic. Further assessed for the study by an experienced clinician. Criteria used not stated

Grossi et al. (2008)	Canada	Cohort study, pain-controlled, linked to treatment trial	20 IBS; 20 nrTMD; 20 rTMD; duration range 3–5 years	Mean age (SD): rTMD=29.4 (9); nrTMD=26.7 (9); IBS=32.9 (1.5), range 15 – 45. Postsecondary education: rTMD=74.3%; nrTMD=50%; IBS=75%; Employed: rTMD=74.3%; nrTMD=58.3%; IBS=50%. High income: rTMD=60.6%; nrTMD=16.7%; IBS=35%	IBS, rTMD and nrTMD groups acted as controls for each other	IBS diagnosis confirmed by physician / based on Rome guidelines. TMD assessed using research diagnostic criteria
Herbert et al. (2018)	USA	Cohort study linked to treatment trial	122 veterans with chronic non-terminal pain. Pain on most days for minimum six months. 45.3% degenerative disc disease, 21.4% OA, 12.8% MSK	Mean age (SD)=52.1 (13.6), range 25 – 89; 96 (82.1%) male; 64 (54.7%) Caucasian; Education: 20 (17.1%) high school or less, 39 (33.3%) some college, 20 (17.1%) community college/trade school, 37 (31.6%) bachelor degree or higher, one not reported	n/a	Diagnosis verified by a study physician based on medical evaluation and medical records. Pain on most days for minimum six months, with average pain severity and inference minimum 4/10 in the past week
Latysheva et al. (2019)	Russia	Cross-sectional, controlled	CM (n=144). Duration Mdn=3, IQR 1, 5 years. Age of Chronic Migraine onset, mean years=36 (range=25-46)	Median age=42.5 (IQR 18, 59); Female/male=132/12; Median education=42.5 years (Interquartile range 12, 15)	Clinical control group with episodic migraines (n=44)	Diagnosis defined by International Classification of Headache Disorders III beta, diagnosis made by a specialist headache neurologist

Oncu et al., (2015)	Turkey	Cross-sectional, controlled	FM (n=96) with a Widespread Pain Index (WPI) score >7 and symptom severity >5 or WPI 3-6 and symptom severity >9; at least three months	Mean age (SD): 32.3 (6.0), range 20-45; Years of education (SD): 10.1 (2.7); All female sample	HCs (n=73); Clinical Controls (n=93) women aged 60-75 with mild cognitive complaints	Diagnosed according to American College of Rheumatology (ACR) 2010 and ACR 1990 criteria
Pidal-Mirandaet al. (2018)	Spain	Cross-sectional, controlled	FM (n=38)	Mean age (SD)=47.71 (9.63), range 28 – 64; all women; Education: 36.8% primary school, 36.8% high school, 26.3 higher studies	HCs (n=33)	Diagnosed by a rheumatologist according to Wolfe 1990 criteria
Radanov et el. (1999)	Switzerland	Cross-sectional	LWS n=21. Persistent head or neck pain and minimum of six months since whiplash injury (mean=26.1, SD=20.7, range 6 - 48 months since injury)	Mean age (SD)=42.2 (8.6), range 20 - 55; 11 (52%) female; 17 (81%) involved in litigation	n/a	Diagnostic or other criteria not explicitly stated but participants were under the care of a Neurology department and underwent assessment during the study
Torkamani et al. (2015)	UK	Cross-sectional	CCH (n=11); duration mean years (SD)=14.64 (11.48)	Mean age (SD)=48.18 (11.02); 82% male; Education mean years (SD)=14.36 (4.2) Age at onset (SD)= 34.55 (10.85)	n/a	International Headache Classification II criteria

CBP=Chronic Back Pain; CP=CP; FM=FM; MSK=Muscoloskeletal; RA=Rheumatoid Arthritis; VP=Visceral Pain; IP=Idiopathic Pain; IBS=Irritable Bowel Syndrome; nrTMD=non-responding Temporomandibular Disorder; rTMD=responding Temporomandibular Disorder; OA=Osteoarthritis; CM=Chronic Migraine; LWS=Late Whiplash Syndrome; CCH=Chronic Cluster Headaches

Table 2. Main Findings

Authors (Year)	WM Measure(s)	Depression Measure	Type of Analysis	Statistical Analyses	Effect Size
Baker et al. (2017)	BRIEF-A WM Subscale. Outcome: Subscale score converted to T-score, then compared against clinical cut-off (>65) based on age-adjusted normative data. Higher score=greater deficit	BDI Outcome: BDI score. Higher score=increased depressive symptoms.	Spearman partial correlation controlling for age and education level	Participants included in analysis (n=39): p<0.05, rho=0.50 Higher WM deficit associated with greater depressive symptoms	Large
Dick et al. (2002)	TEA auditory-verbal WM domain. Outcome: Age-adjusted scaled score. Lower score=greater deficit	HADS-D Outcome: HADS-D score. Higher score=increased depressive symptoms.	Pearson bivariate correlation	Participants included in analysis (n=60): p>0.05 No evidence of an association between WM deficits and depression. *r* not reported	Not reported
Elvemo et al. (2015)	N-back and PVSAT PVSAT Outcome: Sum of correct responses. Lower score=greater deficit. N-back outcome: total correct responses. Lower score=greater deficit.	BDI	Spearman bivariate correlation	Participants included in analysis (n=15): p>0.05 for both analyses N-back and BDI r=0.012 PVSAT and BDI r=-0.39 No evidence of an association between WM deficits and depression when using n-back. Evidence of a relationship between WM deficits and greater depression when using PVSAT, although not statistically significant.	N-Back and BDI=small PVSAT and BDI=medium
Grossi et al. (2008)	CVLT and CCC Outcomes: CVLT - number of correct	BDI	Logistic regression full model	Participants included in analysis (rTMD vs IBS n=56; nrTMD vs IBS n=44):	Small

	responses (CR), word clusters (CL); CCC - total number of correct responses		including study group, income, pain at rest, pain duration, depression, sleep disorders (odds ratio)	rTMD vs IBS: CVLT-CR BDI OR=1.04; CVLT-CL BDI OR=1.04, CCC BDI OR=1.10 nrTMD vs IBS: CVLT-CR BDI OR=1.13 ; CVLT-CL BDI OR=1.14; CCC BDI OR=1.13 None statistically significant as Critical Odds Ratio=2 No evidence of a relationship between WM deficits and depression in any analysis.	
Herbert et al. (2018)	WAIS-IV LNS subtest; CVLT-II. WAIS-IV LNS outcome: age-adjusted scaled score. Lower score=greater deficit	PHQ Outcome: PHQ score. Higher score=increased depressive symptoms	Pearsons bivariate correlation	Participants included in analysis (n=117): LNS and PHQ p> 0.05, r=-0.05 CVLT-II and PHQ p>0.05, r=-0.13 No evidence of a relationship between WM deficits and depression.	Small for both analyses
Latysheva et al. (2019)	RAVLT Outcome: RAVLT total learning score (number of words recalled correctly in five trials). Lower score=greater deficit	HADS-D	Spearman bivariate correlation	Participants included in analysis (n=144) p>0.05, r=0.07 No evidence of a relationship between WM deficits and depression.	Small
Oncu et al., (2015)	STMS Immediate and Delayed recall items Outcome: STMS immediate and delayed recall scores. Lower	BDI	Pearson bivariate correlation	Participants included in analysis (n=86) STMS immediate recall and BDI p=0.437, r=0.109	Small for both analyses

	score=greater deficit			STMS delayed recall and BDI p=0.652, r=-0.063 No evidence of a relationship between WM deficits and depression.	
Pidal-Mirandaet al. (2018)	WAIS-III BSL subtest Outcome: Age-adjusted scaled score. Lower score=greater deficit	BDI	ANCOVA	Participants included in analysis (n=38): WM differences between clinical and HC groups (F(1,59)=5.474, p=0.023, η2=.085) disappeared after depression added as a covariate (F(1,58)=0, p=0.986, η2=0.0) Evidence that higher WM deficits are associated with greater depression symptoms.	n/a
Radanov et el. (1999)	N-back task Outcome: Total correct responses. Lower score=greater deficit	BDI	Spearman bivariate correlation	Participants included in analysis (n=21): p=0.149 (2-tailed) r=-0.3264 Evidence that higher WM deficits are associated with greater depression symptoms, although not statistically significant.	Medium
Torkamani et al. (2015)	CVLT, WSRMF, WAIS-III WMI. Outcomes: WSRMF – Correct responses. WAIS-III WMI – age adjusted WMI index score and scaled scores for individual subtests. Lower scores=greater deficit	BDI	Pearson bivariate correlation	Participants included in analysis (n=11): p>0.05 for correlations between all WM measures and BDI No evidence of a relationship between WM deficits and depression. *r* not reported.	Effect size not reported.

BRIEF-A=Behaviour Rating Inventory of Executive Function; BDI=Beck Depression Inventory; TEA=Test of Everyday Attention; HADS-D=Hospital Anxiety and Depression Scale Depression Subscale; CVLT=California Verbal Learning Test; CCC=Trigram Auditory Memory Task; WAIS-IV=Weschler Adult Intelligence Scale Fourth Edition; LNS=Letter Number Sequencing; PVSAT=Paced Visual Attention Test; PHQ=Patient Health Questionnaire; RAVLT=Rey Auditory Verbal Learning Test; STMS=Short Test of Memory Status; WAIS-III=Weschler Adult Intelligence Scale Third Edition; BSL=Backwards Spatial Localisation; ANCOVA=Analysis of Covariance; WSRMF=Warrington Short Recognition Memory for Faces; WMI=WM Index

Summary of Main Findings

The main findings are summarised in Table 2. Due the fact that several papers used more than one measure of WM, across the 10 papers there were 20 calculations involving regression or correlation between measures of WM and depression. Only one paper reported a significant association between WM and depression (Baker et al., 2018). This was in the hypothesised direction where higher depression scores were associated with increased difficulties with WM, and with a large effect size (rho=0.5). No other analysis found a statistically significant association. However, one paper (Pidal-Miranda et al., 2018) found that when depression was added as a covariate in an ANCOVA model, a previously significant effect of poorer WM performance ($F(1,59)=5.47$, $p=0.023$, $\eta2=0.09$) in the clinical group compared to the HC group disappeared ($F(1,58)=0$, $p=.986$, $\eta2=0.0$). Although the exact association was not reported in this paper, the substantial change in statistical significance and effect size indicates that there was a noteworthy impact of depression on WM in the clinical group as a confounding variable. Among the remaining papers which did not find a statistically significant association, six reported effect sizes covering fourteen separate calculations of the relationship between WM and depression. Two papers, including four calculations of association, did not report effect sizes. Of the six papers which reported effect sizes, four reported negligible effects, incorporating 10 analyses, one reported a medium effect size using one analysis (-0.33), and one study reported one analysis with no effect size and one analysis with a medium effect size (-0.39). The two analyses with medium effect sizes were in the hypothesised direction, with scores on measures of WM decreasing with increased scores on measures of depression.

Subgroup analysis by type of diagnosis

The studies were then divided into three subgroups, based on type of CP diagnosis. In accordance with the International Classification of Diseases, 11th Edition (ICD-11; World Health Organisation, 2019) the

studies included in this review were apportioned to different subgroups based on whether they were considered to be investigating clinical samples with (i) only Chronic Primary Pain (CPP) conditions (ii) Chronic Secondary Pain (CSP) conditions and (iii) those with a combined sample of CPP and CSP conditions. All of the papers included in this study were undertaken before the ICD-11 was published, so did not define participants in terms of CPP or CSP. However, based on the definitions provided above, it was possible in most cases to determine whether the samples recruited would be likely to have met criteria for CPP or CSP from the sample descriptions. Definitions of CPP provided by Nicholas et al. (2019) were also used for further reference during this process. There were six studies investigating CPP only, two investigating CSP only and two investigating a combination. In the two studies investigating a combination, one had a majority of CSP participants (75%) while the other had a 50:50 ratio of CPP and CSP participants. The spread of high-quality studies was similar across subgroups, with CPP-only studies containing two (33%), CSP-only studies containing one (50%) and combined samples containing one (50%). The CSP-only studies contained one low quality paper (50%), while there were four (67%), none (0%) and one (50%) medium quality papers in the CPP, CSP and combined groups respectively. This suggests that the quality of the papers was not a major consideration when interpreting the findings, although the presence of the low-quality paper in the CSP-only papers may be relevant.

Relationship between WM and depression in CPP-only clinical samples

Six studies investigated the association between WM and depression in clinical samples containing only participants who had CPP (Baker et al., 2018; Elvemo et al., 2015; Latysheva et al., 2019; Oncu et al., 2015; Pidal-Miranda et al., 2018; Torkamani et al., 2015). One of the studies found a significant association between WM and depression, where depression scores were positively correlated with WM deficits (Baker et al., 2018). One study comparing WM capacity between clinical and HC groups reported

a notable reduction in statistical significance and effect size when depression was included as a covariate (Pidal-Miranda et al., 2018). This indicated that depression accounted for a proportion of the between-groups WM variance. Two studies found no statistically significant outcomes with negligible or small effect sizes (Latysheva et al., 2019; Oncu et al., 2015). One study found no statistically significant outcomes and did not report effect sizes (Torkamani et al., 2015). One study conducted 2 analyses of associations between WM and depression, neither of which were significant but one had a negligible effect size, while the other had a medium effect size in the hypothesised direction (Elvemo et al., 2015). These findings suggest that in CPP-only samples there is mixed evidence of a relationship between WM and depression, with half of the studies reporting some evidence of an association in the hypothesised direction.

Relationship between WM and depression in CSP-only clinical samples

Two studies investigated the association between WM and depression in clinical samples containing only participants who had CSP (Herbert et al., 2018; Radanov et al., 1999). Neither study found a statistically significant association between WM and depression. One study (Herbert et al., 2018) reported two analyses of different measures of WM with depression, with both producing negligible effect sizes. The other study (Radanov et al., 1999) conducted one correlation analysis between WM and depression, which had a medium effect size in the hypothesised direction. These findings suggest that there is no clear evidence of a relationship between WM and depression in CSP-only samples.

Relationship between WM and depression in clinical samples with a combination of CPP and CSP

Two studies investigated the association between WM and depression in clinical samples containing a combination of participants with CPP and CSP conditions (Dick et al., 2002; Grossi et al., 2008). The composition of the Dick et al. (2002) paper was 50% CPP and 50% CSP participants split across 2 groups. It should be noted that in this study there was a third clinical group but it was not possible to determine whether this group had CPP or CSP from the information available. The composition of the Grossi et al. (2008) paper was 75% CSP and 25% CPP. Neither paper reported a statistically significant relationship between WM and depression. One paper (Dick et al., 2002) did not report an effect size for one correlation computed. The other (Grossi et al., 2008) reported six separate regression outcomes for different WM measures and depression. Interpretation of the reported Odds Ratios according to calculations reported by Chen, Cohen & Chen (2010) indicates that these regression outcomes were not statistically significant and had negligible effect sizes. These papers suggest that in combined groups of CSP and CPP there is no clear evidence of a relationship between WM and depression. Considering one of these papers had a 75% composition of CSP participants, this may suggest that there is further evidence that there is no association between WM and depression in CSP samples particularly.

Subgroup analysis by WM measure type

As previously described, a wide range of measures of WM were used to conduct the 20 analyses of the association between WM and depression across the 10 included studies. It is therefore possible that there were different outcomes for different types of measure. The papers included in this review were divided into subgroups according to whether the measures of WM used (i) auditory-verbal stimuli (ii) visual stimuli, or (iii) were a self-report measure. There were six papers which included an analysis of the association between an auditory-verbal measure of WM and depression, with three (50%) being high quality and three being medium quality (50%). There were four papers which included an analysis of the

association between a visual measure of WM and depression, of which one (25%) was high quality, two (50%) were medium quality and one (25%) was low quality. There was one medium quality paper which included an analysis of the association between a self-report measure of WM and depression. This suggests that the papers reporting analyses involving auditory-verbal measures of WM may be higher quality than the other categories.

Relationship between auditory-verbal WM and depression

Six papers included analyses of the association between WM and depression using WM measures with auditory-verbal stimuli (Dick et al., 2002; Grossi et al., 2008; Herbert et al., 2018; Latysheva et al., 2019; Oncu et al., 2015; Torkamani et al., 2015). None of these papers found a statistically significant association, and four of the studies (Grossi et al., 2008; Herbert et al., 2018; Latysheva et al., 2019; Oncu et al., 2015) reported negligible or small effect sizes (correlations ranging from r=-0.15 to r=0.109 and Odds Ratios ranging from 1.04 to 1.14). Two papers (Dick et al., 2002; Torkamani et al., 2015) did not report an effect size.

Relationship between visual WM and depression

Four papers included analyses of the association between WM and depression using WM measures with visual stimuli (Elvemo et al., 2015; Pidal-Miranda et al., 2018; Radanov et al., 1999; Torkamani et al., 2015). While none of these papers reported a significant association between WM and depression, one paper (Pidal-Miranda et al., 2018) found that a significant difference and medium effect size between the clinical and HC groups on a visual WM measure disappeared after depression was included as a covariate. Two papers (Elvemo et al., 2015; Radanov et al., 1999) reported medium effect sizes (r=-0.39 and r=-0.33 respectively) for the association between visual WM and depression in the hypothesised direction. One paper did not report an effect size (Torkamani et al., 2015).

Relationship between WM using a self-report measure and depression

One paper (Baker et al., 2018) used a self-report measure of WM in its analysis of the relationship between WM and depression. This study found a significant association between WM and depression, with a large effect size, in the hypothesised direction.

Discussion

This systematic review identifies and synthesises literature exploring the relationship between WM and depression in people with a diagnosis of a CP condition. A narrative synthesis of the 10 included studies suggests that overall, WM is not significantly associated with depression in people with a diagnosis of a CP condition. However, when studies were divided into subgroups, there was a less conclusive pattern of results. In particular, the type of CP condition and type of WM measured may provide different outcomes, and statistical power is a further consideration to interpret the findings.

The clearest finding is that nine of the included studies did not find a statistically significant association between WM and depression in samples with CP. One study did find a statistically significant association between WM and depression in a CP sample, which fitted with the original hypothesis of a positive association between WM deficits and heightened depression. These findings overall appear to provide some evidence that there is not a reliable association between these factors in CP. However, the evidence is diminished by insufficient sample sizes in the majority of studies, so should be interpreted cautiously. The one study which did find a statistically significant association was the only study which used a self-report measure of WM. While the measure used in this study was validated, self-report measures are subject to a number of biases and limitations, including introspective ability, ability to interpret the questions, ability and willingness to be honest, and social desirability bias (Devaux & Sassi, 2015). According to Beck's (1987) cognitive model, people with depression are prone to attend to and ruminate on negative information, leading to negative interpretation of information relating to

themselves, disproportionate recall of memories with negative valence and low self-esteem (Disner, Beevers, Haigh & Beck, 2011). Consequently, one interpretation of the positive association between WM and depression when a self-report measure of WM was used is that increased levels of depression may have caused a negatively biased self-perception, poor self-esteem and disproportionate attendance to and recall of instances of WM difficulties. The fact that the other nine studies used less subjective measures may suggest that their findings are more reliable, and they all found no statistically significant association between depression and WM. However, as most studies were underpowered to detect small or medium effect sizes, there is no clear overall conclusion to draw from the 10 studies combined. Considering the evidence that depression and cognitive difficulties including WM are related in other clinical populations (Rose & Ebmeier, 2006), and that there are widespread difficulties with depression and WM in CP, this would suggest the need for more, higher quality and sufficiently powered studies in CP.

Considering the lack of clarity concerning the overall trends, analysis by subtypes of paper might be informative. The following sections will focus on two subdivisions of papers which are pertinent to current conceptualisations of CP and WM. More specifically, comparing papers which consider individuals with a diagnosis of CPP or CSP, as well as comparing papers which use auditory-verbal or visual measures of WM. Subgroup analysis by CP condition type suggested that while there was no evidence of an association between WM and depression in samples of people with a CSP condition only, there was preliminary evidence in studies with samples who had a CPP condition only. In studies which used CPP-only clinical groups, one reported a statistically significant outcome in the hypothesised direction, with a large effect size. Another CPP-only study reported no relevant statistical significance or effect sizes, but did report a notable effect of depression on between-groups differences in WM in an ANCOVA model, in the hypothesised direction. A further study which used a CPP-only clinical sample did not reach statistical significance but reported a medium effect size in the hypothesised direction based

on a relatively small sample. The power calculation for this study did not consider the correlation between WM and depression, but based on the power calculated for this review, it was certainly underpowered to detect a medium effect size at statistical significance. One further study using a CPP-only clinical sample did not report effect sizes, and had the smallest clinical sample of all studies. It is therefore not possible to determine whether the lack of statistical significance reported is due to a low effect size or a small sample. The other two studies which involved CPP-only clinical samples did not find a statistically significant association between WM and depression, reported negligible effect sizes and had relatively large samples. These two studies would have been sufficiently powered to detect a medium effect size, but not a small effect size. A study with 144 participants was considered high quality in our ratings, while one with 96 participants was rated as medium quality. It is possible considering their quality ratings and large sample sizes that these two papers reflect the true picture, and that there is no significant relationship between depression and WM in CPP. However, this conclusion may still be premature based on only two papers, which were still insufficiently powered to detect a small effect size, if one existed. In summary, out of the six studies which used a CPP-only clinical sample, two suggested a relationship between WM and depression, two suggested no relationship, and it was not possible to fully interpret two due to lack of effect size reporting and small sample sizes. The distinction between CPP and CSP in the context of this review may be pertinent. The primary difference between CPP and CSP is that CSP conditions are generally more attributable to a physical trauma than CPP conditions (Treede et al., 2019). CP conditions are caused and maintained by a complex interaction of biological, psychological and social (biopsychosocial) factors (Nicholas et al., 2019), and the distinction between CPP and CSP indicates different biological contributions to the condition. This review suggests that there may be greater – if inconsistent – evidence of a relationship between depression and WM in CPP than CSP. One possible explanation for this might be that while CSP conditions are generally linked to an observable physical trauma, physical changes in CPP are related to less observable changes in the

CNS, including central sensitisation and cortical reorganisation (Ji et al., 2018; Vartiainen et al, 2009). Considering the differential effects on the CNS in CPP and CSP, it may follow that the relationship between processes which are mediated by the brain, such as WM and depression, are more significant in CPP. A further tentative explanation for more evidence of a relationship in CPP between WM and depression is that there may be a greater influence of psychological factors such as depression in CPP within a biopsychosocial conceptualisation, with WM being one area which is affected. While there is no current literature to support either potential explanation, the distinction between CPP and CSP was only established in 2019, meaning that research base is young. Experimental research exploring the prevalence of depression and how it interacts with cognitive functioning between CPP and CSP groups would be informative for determining whether there are differing contributions of biopsychosocial factors.

Another potential interpretation of the findings of this review is that the type of WM measure used produced different patterns of associations between WM and depression. According to Baddeley and Hitch's (1974) seminal model of WM, there are separate visual and auditory-verbal WM processes termed the visuospatial sketchpad and the phonological loop respectively. These processes have been found to be sub-served by different brain regions in imaging studies (Salmon et al., 1996), suggesting an underlying neurological distinction. Considering this, subgroup analysis based on whether visual or auditory-verbal measures of WM were used in the included studies was conducted. The outcome of this analysis found that of the six studies which used an auditory-verbal measure of WM, none reported a statistically significant association between WM and depression, and all reported negligible to small effect sizes or did not report an effect size. While none of the four studies which used a visual, non-verbal measure of WM reported a statistically significant association between WM and depression, three had other statistical indicators of some level or type of relationship, including two with medium effect sizes in the hypothesised direction, and one with a notable contribution of depression to WM

deficits in the CP sample. The two studies which found medium effect sizes had small sample sizes, so it is possible that statistical significance was not reached due to being underpowered rather than a true lack of association. It should also be noted that one of these studies was rated as low quality, meaning that its findings cannot be considered as robust as the other studies in this category. The fourth study which used a visual measure of WM did not report effect sizes, which means that it is not possible to determine whether the failure to reach statistical significance was due to a negligible effect size or being underpowered. The subgroup analysis by type of WM measure suggests that while there is no indication of auditory-verbal WM being associated with depression in CP, there is a majority of papers measuring visual WM where there is some association between WM and depression, although these do not reach statistical significance. While this is based on a small number of papers with unclear statistical outcomes, it raises the possibility that the cognitive and neural processes associated with visual WM processes may be more vulnerable to the impact of depression than auditory-verbal WM in CP.

There is no current literature exploring the interaction of depression with visual and auditory-verbal components of WM in CP differentially, but there is some evidence that brain regions which are associated with depression are also associated with visual, but not auditory-verbal WM. For instance, imaging studies have demonstrated that visual WM tasks activate more extensive bilateral parietal regions than auditory-verbal WM (Na et al., 2000), while depression has been found to be associated with abnormal parietal activity and structure (Zhang, Peng, Sweeney & Gong, 2018). Considering the established role of the parietal lobe in pain perception and modulation (Duncan & Albanese, 2003), there may be neurological factors contributing to an exacerbated relationship between depression and visual WM deficits specifically in CP. There is limited research at the behavioural level concerning the relationship of depression with individual components of WM, but inconsistent associations between depression and the visuospatial sketchpad and phonological loop when dissociated experimentally in

non-CP populations have been found (Channon, Baker & Robertson, 1993; Christopher & Macdonald, 2005). These studies either found that depression did not have a relationship with either the visuospatial sketchpad or the phonological loop, or found that it was negatively associated with performance on both. Further research by Thompson et al. (2007) found that scores on a measure of WM involving the visuospatial sketchpad was significantly poorer for a bipolar disorder sample than a HC group, while scores on a measure of WM involving the phonological loop were not. This provides some evidence that a depressive condition can be associated with visuospatial sketchpad but not phonological loop difficulties, although there is no literature investigating this distinction in CP, with or without depression as a factor. This gap in knowledge is highlighted by the findings of this review, which warrants empirical investigation. It should also be noted that none of the studies investigated WM using other modalities, most notably somatosensory WM or auditory WM without a verbal component. These types of WM involve other cognitive and neural processes and the relationship of these processes to depression remains unknown.

Limitations

This review used a narrative synthesis approach, meaning further research is needed to provide more concrete evidence for or against the suggested outcomes hypotheses. There was also a relatively low number of papers included, which means that the findings, particularly those involving subgroups of only one or two papers, should be treated with caution. The low number of studies also means that papers with low quality, or which did not report full statistical outcomes, which accounted for 40% of the total, reduced the ability to interpret the findings with confidence even further. Another limitation of this review is that all of the studies were observational, meaning that there is no indication of causality. As such, any findings suggesting some type of relationship between WM and depression do not provide an indication of whether depression affects WM, vice versa, or if there is bidirectional causality.

Clinical Implications

This review provides preliminary evidence that the relationship between WM deficits and depression may be more pronounced in people with a CPP condition than a CSP condition. While this is far from a firm conclusion, none of the papers investigating CSP-only samples, or samples with a 50% or higher composition of CSP participants found any association, while a third of CPP-only papers found some statistical evidence of a relationship between WM and depression. Half of CPP-only papers which could be fully interpreted found evidence of an association between WM and depression. As such, there appears to be some difference between CPP and CSP groups. This raises the interesting possibility that there may be more of an interplay of cognitive and psychological factors in CPP than CSP, and possibly a greater contribution of these factors to the development and maintenance of CPP conditions. More research is clearly needed to determine whether this hypothesis stands up to scrutiny, but if such a difference is found it could contribute to more tailored interventions for people with CPP and CSP conditions.

Among the core competencies for Clinical Psychologists in the UK are assessment and formulation, where relevant information is gathered and then a clinical hypothesis is developed for presenting problems (DCP, 2011). This process is based on theoretical understandings of clinical populations and presenting issues, drawn from the evidence base (Johnstone & Dallos, 2013). Bespoke interventions are then developed to suit the needs of clients arising from the formulation, so a refined evidence base is crucial for underpinning precise and relevant interventions (DCP, 2011). If the findings of this literature review are corroborated by empirical studies, the knowledge gained could contribute to more efficient and personalised formulations, in which explanations for depression and WM deficits could be understood as maintaining factors for each other in some circumstances but not others. For instance,

clinicians could be mindful, when assessing clients with CPP, that reported difficulties with depression may be related to WM deficits, or vice versa, but when these issues are found during assessment of clients with CSP, they are less likely to be related. This would guide the assessment; for instance, if a clinician encounters a client with CPP who reports depression, an assessment of WM may be useful for conceptualising their difficulties, while for someone with CSP reporting depression, it may not be. Furthermore, an assessment of visual WM for a client with CPP reporting depression would be more likely to identify a WM deficit which is related to their low mood than an auditory-verbal assessment. Based on the assessment, if difficulties with both depression and visual WM were found for a client with CPP, the case formulation would hypothesise that these difficulties contributed to each other, which is an important step towards developing an appropriate intervention. Similarly, if a clinician assesses a client with CPP who reports WM difficulties, it may also be more relevant to gather information about depression and include that in the formulation, than for someone with CSP who reports WM difficulties.

Formulations based on refined theoretical conceptual understandings of specific clinical populations identify potential interventions which can interrupt negative patterns which maintain presenting problems (Johnstone & Dallos, 2013). A strength of psychological formulation and intervention is that it is unique to each individual (Johnstone & Dallos, 2013) so permutations of this knowledge would be context-dependent. However, an example of such an intervention for a client with CPP who reports difficulties with WM and also presents with symptoms of depression, might involve clinicians treating depression using evidence-based psychosocial interventions such as behavioural activation (NICE, 2009), alongside neuropsychological interventions targeting visual WM, such as cognitive remediation (Bell et al., 2003). This would maximise the chances of improvements in both WM and mood domains by not only targeting each individually but recognising that they may be maintaining factors for each other, requiring an additional breaking of a maintenance cycle. Without such an approach, it is possible that attempts to remediate WM deficits would be undermined or degraded over time by lingering untreated

depressive symptoms, or vice versa. A joined-up approach to treatment could therefore make service provision more efficient and client-centred, in line with NHS core values (Department of Health and Social Care, 2021).

In support of this combined approach to treatment, there is already evidence that in other clinical populations interventions targeting WM can improve depressive symptoms. Some cognitive remediation interventions targeting WM capacity directly, such as Cognitive Control Training, have shown positive outcomes in terms of reduced rumination and depression (Siegle et al., 2007), and these effects can be enhanced and elongated by transcranial Direct Current Stimulation (tDCS), which contributes additional neural stimulation (Brunoni et al., 2014). Joplin et al. (2020) additionally demonstrated that training participants who met diagnostic criteria for major depressive disorder to remove negative information from WM led to reduced depressive symptoms including rumination, negative intrusive thoughts and stress responses. Although there is currently no literature concerning the effect of interventions targeting depression on WM capacity, this latter study raises the intriguing possibility that interventions such as Cognitive Behavioural Therapy which aim to reduce negative ruminations and intrusive negative thoughts (Gautam et al., 2020), might benefit WM by decreasing the load of information held therein. In practice, while each case would need individual consideration, an example of an intervention based on the established literature around supporting difficulties with WM and depression, might involve: (i) compensatory strategies such as writing information in notebooks or diaries to support visual WM; (ii) direct cognitive rehabilitation such as Cognitive Control Training to target WM capacity directly; (iii) potentially direct neural stimulation using tDCS to enhance cognitive rehabilitation, where possible; (iv) psychosocial interventions such as CBT or behavioural activation to reduce depressive symptoms. While these are all established interventions, this study has provided preliminary evidence to suggest that combining these types of intervention could optimise outcomes in terms of both depression and WM.

Additionally, this study suggests that this combination of approaches could be more beneficial for those with CPP than CSP, and that visual working memory may be a more fruitful avenue for intervention than auditory-verbal. As such, it has contributed valuable new lines of enquiry which could lead to more efficient service provision.

Rehabilitation for WM is not commonly undertaken in CP treatment settings and there is no research on the impact of WM rehabilitation on wider functioning or quality of life in CP (Baker et al., 2017). However, following mild traumatic brain injury, it has been demonstrated that WM training can improve the ability to undertake activities of daily living and return to employment (Vallat-Azouvi et al., 2009). Psychosocial interventions targeting depression in CP have been shown to decrease difficulties with mood, as well as increase activity levels and positive coping (Holmes et al., 2007). As such, finding ways to enhance and increase treatment options for WM and depression can be expected to deliver further benefits for the functioning and emotional wellbeing of people living with CP. The new understandings of the potential for WM difficulties and depression to maintain one another in CPP found in this review, and the implication for adopting treatment approaches which target WM and depression concurrently, therefore have potential for improving clients' functional and emotional outcomes. As the potential relationship between WM and depression in CPP is a new finding, in order to integrate this new knowledge and implications for treatment into pain management services, training may be beneficial. Recently, there have been calls for greater training and a defined training curriculum in the specialty of pain psychology (Wandner et al., 2019). This specialist training needs to integrate knowledge and competencies specific to the chronic pain population, differentiated from other clinical psychology specialties (Cox et al., 2013). If verified through empirical studies, the new insights from the current review could be incorporated into this specialist training, as it demonstrates a profile of cognitive and

emotional characteristics and interactions which are unique to CP and are therefore important for specialist pain psychologists to be aware of.

Conclusions

This is the first review of the literature concerning the relationship between WM and depression in CP. Overall the findings suggest that there is no statistically significant relationship between WM and depression in CP populations that cannot be explained by other factors such as self-report biases. However, subgroup analysis by type of CP diagnosis and type of WM measured suggests that there may be some benefit from conceptualising the relationship between WM and depression differentially according to these categories. There was incomplete data and one low quality study, which means that the interpretation of these findings and their wider generalisability is limited. Further empirical research is needed to determine whether these indications could be informative for pain management services. More high quality research with sample sizes sufficient to detect small to medium effect sizes is needed to confirm or refute the preliminary findings of this review. The issue of causality between depression and WM, if any exists, has not been researched, and longitudinal cohot studies appear to be the most practicable and ethical way to elucidate this question.

References

Antepohl, W., Kiviloog, L., Andersson, J., & Gerdle, B. (2003). Cognitive impairment in patients with chronic whiplash-associated disorder--a matched control study. *NeuroRehabilitation*, *18*(4), 307–315.

Baddeley, A.D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8,* 47-89.

Baddeley, A.D. (2010). Working memory. *Curr Biol. 23*(4), 136-140. https://doi.org/10.1016/j.cub.2009.12.014

Baker, K. S., Georgiou-Karistianis, N., Gibson, S. J., & Giummarra, M. J. (2017). Optimizing Cognitive Function in Persons With Chronic Pain. *The Clinical Journal of Pain*, *33*(5), 462–472. https://doi.org/10.1097/AJP.0000000000000423

Baker, K. S., Gibson, S. J., Georgiou-Karistianis, N., & Giummarra, M. J. (2018). Relationship between self-reported cognitive difficulties, objective neuropsychological test performance and psychological distress in chronic pain. *European Journal of Pain (London, England)*, *22*(3), 601–613. https://doi.org/10.1002/ejp.1151

Beck, A. T. (1987). Cognitive models of depression. *J. Cogn. Psychother. 1*, 5–37.

Bell, M., Bryson, G., & Wexler, B. E. (2003). Cognitive remediation of working memory deficits: Durability of training effects in severely impaired and less severely impaired schizophrenia. *Acta Psychiatrica Scandinavica, 108*(2), 101–109. https://doi.org/10.1034/j.1600-0447.2003.00090.x

Benjamin, S., Morris, S., McBeth, J., Macfarlane, G. J., & Silman, A. J. (2000). The association between chronic widespread pain and mental disorder: a population-based study. *Arthritis and Rheumatism*, *43*(3), 561–567. https://doi.org/10.1002/1529-0131(200003)43:3<561::AID-ANR12>3.0.CO;2-O

Berryman, C., Stanton, T. R., Jane Bowering, K., Tabor, A., McFarlane, A., & Lorimer Moseley, G. (2013). Evidence for working memory deficits in chronic pain: a systematic review and meta-analysis. *Pain*, *154*(8), 1181–1196. https://doi.org/10.1016/j.pain.2013.03.002

Boland, A., Cherry, M.G., & Dickson, R. (2017). *Doing a systematic review: a student's guide.* London: Sage.

Brunoni, A. R., Boggio, P. S., De Raedt, R., Benseñor, I. M., Lotufo, P. A., Namur, V., Valiengo, L. C., & Vanderhasselt, M. A. (2014). Cognitive control therapy and transcranial direct current stimulation for depression: a randomized, double-blinded, controlled trial. *Journal of Affective Disorders*, *162*, 43–49. https://doi.org/10.1016/j.jad.2014.03.026

Bushnell, M. C., Ceko, M., & Low, L. A. (2013). Cognitive and emotional control of pain and its disruption in CP. *Nature Reviews. Neuroscience*, *14*(7), 502–511. https://doi.org/10.1038/nrn3516

Channon, S., Baker, J. E., & Robertson, M. M. (1993). Working memory in clinical depression: an experimental study. *Psychological Medicine*, *23*(1), 87–91. https://doi.org/10.1017/s0033291700038873

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation, 39*(4), 860-864, https://doi.org/10.1080/03610911003650383

Christopher, G., & MacDonald, J. (2005). The impact of clinical depression on working memory. *Cogn Neuropsychiatry*, *10*(5), 379-99. https://doi.org/10.1080/13546800444000128

Cox, D. R., Cox, R. H., & Caplan, B. (2013). *Specialty competencies in rehabilitation psychology.* New York, NY: Oxford University Press. http://dx.doi.org/10.1093/med:psych/9780195389241.001.0001

Crombez, G., Van Damme, S., & Eccleston, C. (2005). Hypervigilance to pain: an experimental and clinical analysis. *Pain*, *116*(1-2), 4–7. https://doi.org/10.1016/j.pain.2005.03.035

Department of Health and Social Care (2021). *Guidance: The NHS constitution for England.* Retrieved from https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england

Devaux, M., & Sassi, F. (2016). Social disparities in hazardous alcohol use: self-report bias may lead to incorrect estimates. *European Journal of Public Health*, *26*(1), 129–134. https://doi.org/10.1093/eurpub/ckv190

Dick, B., Eccleston, C., & Crombez, G. (2002). Attentional functioning in fibromaylgia, rheumatoid arthritis, and musculoskeletal pain patients. *Arthritis and Rheumatism*, *47*(6), 639–644. https://doi.org/10.1002/art.10800

Dick B.D., & Rashiq S. (2007). Disruption of attention and wrking memory traces in individuals with chronic pain. *Anesthesia and Analgesia*, *104*(5), 1223-9. https://doi.org/10.1213/01.ane.0000263280.49786.f5

Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews. Neuroscience*, *12*(8), 467–477. https://doi.org/10.1038/nrn3027

Division of Clinical Psychology (2011). *Good practice guidelines on the use of psychological formulation.* Retrieved from http://www.sisdca.it/public/pdf/DCP-Guidelines-for-Formulation-2011.pdf

Duncan, G. H., & Albanese, M. C. (2003). Is there a role for the parietal lobes in the perception of pain?. *Advances in Neurology*, *93*, 69–86.

Dworkin, R.H., Turk, D.C., Farrar, J.T., Haythornthwaite, J.A., Jensen, M.P., Katz, N.P., Kerns, R.D., Stucki, G., Allen, R.R., Bellamy, N., Carr, D.B., Chandler, J., Cowan, P., Dionne, R., Galer, B.S., Hertz, S., Jadad, A.R., Kramer, L.D., Manning, D.C., Martin, S., McCormick, C.G., McDermott, M.P., McGrath, P., Quessy, S., Rappaport, B.A., Robbins, W., Robinson, J.P., Rothman, M., Royal, M.A., Simon, L., Stauffer, J.W., Stein, W., Tollet, JJ., Wernicke, J., & Witter, J. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain, 113,* 9-19.

Elvemo, N. A., Landrø, N. I., Borchgrevink, P. C., & Håberg, A. K. (2015). A particular effect of sleep, but not pain or depression, on the blood-oxygen-level dependent response during working memory tasks in patients with chronic pain. *Journal of Pain Research*, *8*, 335–346. https://doi.org/10.2147/JPR.S83486

Fitzcharles, M. A., Cohen, S. P., Clauw, D. J., Littlejohn, G., Usui, C., & Häuser, W. (2021). Nociplastic pain: towards an understanding of prevalent pain conditions. *Lancet (London, England)*, *397*(10289), 2098–2110. https://doi.org/10.1016/S0140-6736(21)00392-5

Gautam, M., Tripathi, A., Deshmukh, D., & Gaur, M. (2020). Cognitive behavioral therapy for depression. *Indian Journal of Psychiatry*, *62*(Suppl 2), S223–S229. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_772_19

Gelonch, O., Garolera, M., Valls, J., Rosselló, L., & Pifarré, J. (2017). Cognitive complaints in women with fibromyalgia: Are they due to depression or to objective cognitive dysfunction?. *Journal of Clinical and Experimental Neuropsychology*, *39*(10), 1013–1025. https://doi.org/10.1080/13803395.2017.1301391

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: current status and future directions. *Annual Review of Clinical Psychology*, *6*, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Grossi, M. L., Goldberg, M. B., Locker, D., & Tenenbaum, H. C. (2008). Irritable bowel syndrome patients versus responding and nonresponding temporomandibular disorder patients: a neuropsychologic profile comparative study. *The International Journal of Prosthodontics*, *21*(3), 201–209.

Hart, R. P., Wade, J. B., & Martelli, M. F. (2003). Cognitive impairment in patients with chronic pain: the significance of stress. *Current Pain and Headache Reports*, *7*(2), 116–126. https://doi.org/10.1007/s11916-003-0021-5

Herbert, M. S., Afari, N., Robinson, J. B., Listvinsky, A., Bondi, M. W., & Wetherell, J. L. (2018). Neuropsychological functioning and treatment outcomes in acceptance and commitment therapy for chronic pain. *The Journal of Pain*, *19*(8), 852–861. https://doi.org/10.1016/j.jpain.2018.02.008

Holmes, A., Christelis, N., & Arnold, C. (2013). Depression and chronic pain. *The Medical Journal of Australia*, *199*(S6), S17–S20. https://doi.org/10.5694/mja12.10589

Jamison, R. N., Sbrocco, T., & Parris, W. (1989). The influence of physical and psychosocial factors on accuracy of memory for pain in chronic pain patients. *Pain*, *37*(3), 289–294. https://doi.org/10.1016/0304-3959(89)90193-0

Ji, R. R., Nackley, A., Huh, Y., Terrando, N., & Maixner, W. (2018). Neuroinflammation and central sensitization in chronic and widespread Pain. *Anesthesiology*, *129*(2), 343–366. https://doi.org/10.1097/ALN.0000000000002130

Johnstone, L., & Dallas, R. (2014). *Formulation in psychology and psychotherapy: Making sense of people's problems (second edition).* Abingdon, UK: Routledge.

Jopling, E., Gotlib, I. H., & LeMoult, J. (2020). Effects of working memory training on cognitive, affective, and biological responses to stress in major depression: A novel cognitive bias modification protocol. *Journal of Affective Disorders*, *265*, 45–51. https://doi.org/10.1016/j.jad.2020.01.007

Legrain, V., Iannetti, G. D., Plaghki, L., & Mouraux, A. (2011). The pain matrix reloaded: a salience detection system for the body. *Progress in Neurobiology*, *93*(1), 111–124. https://doi.org/10.1016/j.pneurobio.2010.10.005

McCracken, L. M., Spertus, I. L., Janeck, A. S., Sinclair, D., & Wetzel, F. T. (1999). Behavioral dimensions of adjustment in persons with chronic pain: pain-related anxiety and acceptance. *Pain*, *80*(1-2), 283–289. https://doi.org/10.1016/s0304-3959(98)00219-x

Mitchell, A., Kemp, S., Benito-León, J., & Reuber, M. (2010). The influence of cognitive impairment on health-related quality of life in neurological disease. *Acta Neuropsychiatrica, 22*(1), 2-13. https://doi.org/10.1111/j.1601-5215.2009.00439.x

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339,* b2535. https://doi.org/10.1136/bmj.b2535

Molton, I. R., Graham, C., Stoelb, B. L., & Jensen, M. P. (2007). Current psychological approaches to the management of chronic pain. *Current Opinion in Anaesthesiology*, *20*(5), 485–489. https://doi.org/10.1097/ACO.0b013e3282ef6b40

Na, D.G., Ryu, J.W., Byun, H.S., Choi, D.S., Lee, E.J., Chung, W.I., Cho, J.M., & Han, B.K. (2000). Functional MR Imaging of WM in the Human Brain. *Korean J Radiol., 2000 1*(1), 19-24. https://doi.org/10.3348/kjr.2000.1.1.19

National Health Service (2019). *Overview: Arthritis.* Retrieved from https://www.nhs.uk/conditions/arthritis/

National Institute for Health and Care Excellence (2009). *Depression in adults: recognition and management.* Retrieved from https://www.nice.org.uk/guidance/cg90/ifp/chapter/treatments-for-mild-to-moderate-depression

National Institute for Health and Care Excellence (2021). *Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain.* Retrieved from https://www.nice.org.uk/guidance/ng193

Nicholas, M., Vlaeyen, J., Rief, W., Barke, A., Aziz, Q., Benoliel, R., Cohen, M., Evers, S., Giamberardino, M. A., Goebel, A., Korwisi, B., Perrot, S., Svensson, P., Wang, S. J., Treede, R. D., & IASP Taskforce for the Classification of CP (2019). The IASP classification of chronic pain for ICD-11: chronic primary pain. *Pain*, *160*(1), 28–37. https://doi.org/10.1097/j.pain.0000000000001390

Oncu, J., Iliser, R., & Kuran, B. (2015). Cognitive complaints in patients with fibromyalgia versus older-aged women: Case control study. *Journal of Psychiatry, 18*(2), 1-6.

Pacheco-Unguetti, A. P., Acosta, A., Callejas, A., & Lupiáñez, J. (2010). Attention and anxiety: different attentional functioning under state and trait anxiety. *Psychological Science*, *21*(2), 298–304. https://doi.org/10.1177/0956797609359624

Papazacharias, A., & Nardini, M. (2012). The relationship between depression and cognitive deficits. *Psychiatria Danubina*, *24 Suppl 1*, S179–S182.

Peveler, R., Katona, C., Wessely, S., & Dowrick, C. (2006). Painful symptoms in depression: under-recognised and under-treated?. *The British Journal of Psychiatry : The Journal of Mental Science*, *188*, 202–203. https://doi.org/10.1192/bjp.bp.105.012963

Pidal-Miranda, M., González-Villar, A. J., Carrillo-de-la-Peña, M. T., Andrade, E., & Rodríguez-Salgado, D. (2018). Broad cognitive complaints but subtle objective working memory impairment in FM patients. *PeerJ*, *6*, e5907. https://doi.org/10.7717/peerj.5907

Radanov, B. P., Bicik, I., Dvorak, J., Antinnes, J., von Schulthess, G. K., & Buck, A. (1999). Relation between neuropsychological and neuroimaging findings in patients with late whiplash syndrome.

Journal of Neurology, Neurosurgery, and Psychiatry, *66*(4), 485–489. https://doi.org/10.1136/jnnp.66.4.485

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–2040. https://doi.org/10.1017/S0033291713002535

Rose, E. J., & Ebmeier, K. P. (2006). Pattern of impaired working memory during major depression. *Journal of Affective Disorders*, *90*(2-3), 149–161. https://doi.org/10.1016/j.jad.2005.11.003

Sagheer, M. A., Khan, M. F., & Sharif, S. (2013). Association between chronic low back pain, anxiety and depression in patients at a tertiary care centre. *The Journal of the Pakistan Medical Association*, *63*(6), 688–690.

Salmon, E., Van der Linden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., Luxen, A., & Franck, G. (1996). Regional brain activity during working memory tasks. *Brain : A Journal of Neurology*, *119*(5), 1617–1625. https://doi.org/10.1093/brain/119.5.1617

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84*(3), 328-364. https://doi.org/10.3102/0034654313500826

Schnurr, R. F., & MacDonald, M. R. (1995). Memory complaints in chronic pain. *The Clinical Journal of Pain*, *11*(2), 103–111. https://doi.org/10.1097/00002508-199506000-00004

Sharif, S., Saleem, A., Koumadoraki, E., Jarvis, S., Madouros, N., & Khan, S. (2021). Headache - a window to dementia: An unexpected twist. *Cureus*, *13*(2), e13398. https://doi.org/10.7759/cureus.13398

Sheline, Y. I., Wang, P. W., Gado, M. H., Csernansky, J. G., & Vannier, M. W. (1996). Hippocampal atrophy in recurrent major depression. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(9), 3908–3913. https://doi.org/10.1073/pnas.93.9.3908

Siegle, G. J., Ghinassi, F., & Thase, M. E. (2007). Neurobehavioral therapies in the 21st century: Summary of an emerging field and an extended example of cognitive control training for depression. *Cognitive Therapy and Research, 31*(2), 235–262. https://doi.org/10.1007/s10608-006-9118-6

Silver, J. M., McAllister, T. W., & Arciniegas, D. B. (2009). Depression and cognitive complaints following mild traumatic brain injury. *The American Journal of Psychiatry*, *166*(6), 653–661. https://doi.org/10.1176/appi.ajp.2009.08111676

Sturgeon, J. A., Dixon, E. A., Darnall, B. D., & Mackey, S. C. (2015). Contributions of physical function and satisfaction with social roles to emotional distress in CP: a Collaborative Health Outcomes Information Registry (CHOIR) study. *Pain*, *156*(12), 2627–2633. https://doi.org/10.1097/j.pain.0000000000000313

Thompson, J. M., Hamilton, C. J., Gray, J. M., Quinn, J. G., Mackin, P., Young, A. H., & Ferrier, I. N. (2006). Executive and visuospatial sketchpad resources in euthymic bipolar disorder: Implications for visuospatial working memory architecture. *Memory (Hove, England)*, *14*(4), 437–451. https://doi.org/10.1080/09658210500464293

Torkamani, M., Ernst, L., Cheung, L. S., Lambru, G., Matharu, M., & Jahanshahi, M. (2015). The neuropsychology of cluster headache: cognition, mood, disability, and quality of life of patients with chronic and episodic cluster headache. *Headache*, *55*(2), 287–300. https://doi.org/10.1111/head.12486

Treede, R., Rief, W., Barke, A., Aziz, Q., Bennett, M.I., Benoliel, R., Cohen, M., Evers, S., Finnerup, N.B., First, M.B., Giamberardino, M.A., Kaasa, S., Korwisi, B., Kosek, E., Lavande'homme, P., Nicholas, M., Perrot, S., Scholz, J., Schug, S., Smith, B.H., Svensson, P., Vlaeyen, J.W.S., & Wang, S. (2019). Chronic pain as a symptom or a disease: the IASP Classification of chronic pain for the International Classification of Diseases (ICD-11). *Pain, 160*(1), 19-27.

Vallat-Azouvi, C., Pradat-Diehl, P., & Azouvi, P. (2009). Rehabilitation of the central executive of working memory after severe traumatic brain injury: Two single-case studies. *Brain Injury*, *23*(6), 585–594. https://doi.org/10.1080/02699050902970711

Varni, J.W., Rapoff, M.A., Waldron, S.A., Gragg, R.A., Bernstein, B.H., & Linsley, C.B. (1996). Chronic pain and emotional distress in children and adolescents. *Journal of Developmental and Behavioral Pediatrics,17*(3),154-161.

Vartiainen, N., Kirveskari, E., Kallio-Laine, K., Kalso, E., & Forss, N. (2009). Cortical reorganization in primary somatosensory cortex in patients with unilateral CP. *The Journal of Pain*, *10*(8), 854–859. https://doi.org/10.1016/j.jpain.2009.02.006

Veinante, P., Yalcin, I., & Barrot, M. (2013). The amygdala between sensation and affect: a role in pain. *Journal of Molecular Psychiatry*, *1*(1), 9. https://doi.org/10.1186/2049-9256-1-9

Wells, G. A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Loeb, M. (2014). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis.* Retrieved from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Wandner, L. D., Prasad, R., Ramezani, A., Malcore, S. A., & Kerns, R. D. (2019). Core competencies for the emerging specialty of pain psychology. *The American Psychologist*, *74*(4), 432–444. https://doi.org/10.1037/amp0000330

Wu, Y. L., Huang, C. J., Fang, S. C., Ko, L. H., & Tsai, P. S. (2018). Cognitive impairment in fibromyalgia: A meta-analysis of case-control studies. *Psychosomatic Medicine*, *80*(5), 432–438. https://doi.org/10.1097/PSY.0000000000000575

Zhang, F. F., Peng, W., Sweeney, J. A., Jia, Z. Y., & Gong, Q. Y. (2018). Brain structure alterations in depression: Psychoradiological evidence. *CNS Neuroscience & Therapeutics*, *24*(11), 994–1003. https://doi.org/10.1111/cns.12835

Chapter 2: Empirical Paper

Towards a neuropsychological profile of chronic pain: Development and validation of a novel measure of auditory working memory deficits in fibromyalgia.

Word count (excluding references): 14,151

Paper prepared for submission to Frontiers in Pain Research journal for peer review. Author guidelines are in Appendix E.

Abstract

Background: Working Memory deficits are frequently reported and demonstrated through formal assessment in people with a diagnosis of fibromyalgia, but the precise characterisation of these difficulties remains vague. At present, measures of Working Memory only cover a limited range of timescales and sensory modalities, restricting research into the topic and assessment in clinical settings.

Objective: This study aimed primarily to develop and provide evidence for the population and construct validity of a measure of Auditory Working Memory on a timescale of 10s of seconds to minutes by assessing reproducibility across clinical and control groups. The auditory sensory modality was chosen as most practical in the context of the COVID-19 pandemic, allowing home-based online task administration. Its secondary aims were to provide pilot data concerning differences between participants with fibromyalgia and healthy controls on measures of general and Auditory Working Memory; and to provide preliminary evidence for any relationships between clinical and experimental measures of Working Memory, and between general Working Memory and Auditory Working Memory capacity.

Methods: A newly developed measure of Auditory Working Memory was administered to 17 participants with fibromyalgia and 17 healthy controls, aged between 18-64 years old. Fifteen of each group were included in the final analysis. This novel measure manipulated the exposure to different probabilities of auditory stimulus parameters across different blocks in a within-subjects design, allowing to quantify top-down influences (prior experience) on parametric Auditory Working Memory. Accuracy and reaction times were analysed to validate the measure across the two groups separately, and both combined. It was hypothesised that increased probabilities would decrease top-down influences on parametric Auditory Working Memory, while decreased probabilities would lead to an increase in this effect, which would be consistent across groups, providing evidence of population and

construct validity. A partial correlation between the novel measure and existing Working Memory tests was also hypothesised, providing evidence towards criterion validity. Preliminary evidence of Auditory Working Memory deficits for participants with fibromyalgia, and a positive correlation between clinical and experimental measures of Working Memory, which would provide evidence towards concurrent validity of the experimental measure, were further hypotheses.

Results: Consistent outcomes regarding Auditory Working Memory were found across the separate and combined groups in terms of effect sizes and means, suggesting population validity, although statistical significance was only reached for the combined group. As expected, manipulating the probability of auditory stimulus parameters across blocks biased participants' judgements towards their average characteristics, thereby supporting construct validity of the measure, although this was not the case for all probability distributions. Evidence of criterion validity was not found, which may reflect the different constructs being measured in the novel and existing measures. Further analysis found a significant association between clinical and experimental measures of Working Memory with a large effect size, suggesting concurrent validity for the experimental measure, and subtle evidence of poorer performance on measures of Working Memory and Auditory Working Memory by participants with fibromyalgia than healthy controls.

Conclusions: This study has provided evidence of population and construct validity for a measure of Auditory Working Memory on a scale of 10s of seconds to minutes, demonstrating its ability to elicit reproducible top-down effects on working memory, and supporting its utility for further research into Working Memory deficits in fibromyalgia and ultimately for assessment in clinical settings. Some unexpected findings which contradicted the hypotheses, particularly that there was no evidence of an impact of top-down Working Memory on parametric Working Memory in one experimental condition, can be used to inform further refinement of this measurement tool.

Keywords: Fibromyalgia, Working Memory, Auditory, Measure Validation, Within-Subjects

Introduction

Fibromyalgia (FM) is one of the most common Chronic Primary Pain (CPP) conditions, with approximately one in 20 people in the UK affected to some degree (NHS, 2019). CPP conditions such as FM carry enormous personal, social and economic burdens, including physical disability, mental health comorbidity, absenteeism, increased contact and strain on healthcare services and finances, and unemployment (Phillips, 2009; Pain Proposing Steering Committee, 2010). Despite its prevalence and impact, current understanding of the causes of FM is poor and consequently approaches to treatment lack efficacy (Pain Proposing Steering Committee, 2010; Abdel Shahee et al., 2016). It is not known though what factors are causally related to poor functional outcomes in FM.

Mild cognitive deficits such as those found in CP have been found in other clinical populations to be related to reduced functionality and quality of life, for example in samples with mild cognitive impairment and mild dementia (Stites, Harkins, Rubright & Karlawish, 2019). While a range of cognitive domains have been found to be impaired in FM, including attention, free recall, spatial memory and executive functions (Bertolucci & De Oliveira, 2013), working memory (WM) deficits are a cardinal feature of the cognitive dysfunction in this population (Ambrose et al., 2012). A previous meta-analysis by Berryman et al. (2013) found consistent evidence of cognitive deficits, particularly WM deficits, in CP conditions. Forty percent of the 215 included papers evaluated WM in FM. WM refers to the ability to hold information in mind for short periods, ranging from milliseconds to minutes, and mentally manipulate this information to complete activities in vivo (Bjorkdahl et al., 2013). It is a crucial skill for achieving daily goals, so deficits in WM can significantly impact on day-to-day functioning (Garcia-Alvarez et al., 2019). WM operates on a range of timescales and with different levels of abstraction and control, and different studies have also shown that it occurs with respect to any sensory modality, linked

to differential involvement of multiple brain regions (Fassihi et al., 2014; Fuster & Bressler, 2012). Deficits at either basic or more complex stages of this process – where information is either simply held in mind, or a task involves secondary processing components requiring more sophisticated cognitive functions – can hinder goal-oriented activity (Conway et al., 2003). WM can be considered to be related to both the capacity to hold information in mind, and also capacity to maintain attention on that information and rehearse it; these distinct aspects of WM require different timescales (Tetzlaff et al., 2012). As such, WM is a multifaceted phenomenon requiring consideration across a wide spectrum of complexity and timescales.

However, the precise characterisation of WM difficulties found in FM is still unclear. A growing body of evidence suggests that people with CP have deficits on tasks linked to the Posterior Parietal Cortex, a brain region associated with WM (Cohen et al., 2013) and behavioural assessments of WM in CP, including FM, have frequently found deficits (Berryman et al. 2013). The reasons why we might find these deficits in FM include: (i) excessive nociceptive neural activity diverting cognitive resources from other processes, including WM (Berryman et al., 2013); (ii) hypervigilance to bodily sensations, including salient nociceptive inputs, in CP directing attention away from other cognitive processes, including WM (Legrain et al., 2011). This phenomenon is exacerbated by fear and catastrophizing about pain, which is often found in CP (Crombez et al., 1998); (iii) Over-attendance to specific body areas because of hypervigilance to somatosensory stimuli or in search of pain relief leading to stronger neural pain responses to nociceptive stimuli in these regions (Legrain et al., 2009). This over-estimation of irrelevant nociceptive stimuli restricts other information from the environment accessing WM (Legrain et al., 2009); (iv) persistent pain disrupts brain functioning across domains not specific to pain processing. This includes the default mode network (DMN), which provides a balanced baseline resting brain state, essential for helping the brain to maintain information needed for responding to and predicting environmental demands. As such, disrupted DMN functioning in CP undermines WM processes (Baliki et

al., 2008). It has previously been found that cognitive ability including WM may predict the development of CP and outcome in terms of the experience of pain six- and twelve-months post-surgery for patients who previously did not have CP (Attal et al., 2014). This raises the question of whether WM might predict clinical outcomes in FM, and which timescales, levels or sensory modalities of WM could be relevant.

In non-CP studies there have been findings of neural activity in different brain regions depending on timescale of a WM task, ranging from sub-second integration of information (Jadhav & Feldman, 2010) to maintenance of information in WM over seconds to minutes (Romo & Salinas, 2003). Additionally, Baddeley's (2000) updated model of WM proposes that WM is subdivided into four subsystems, with differing timescales of WM involved. This model drew on a proposed 'long-term WM' (Ericsson & Klintsch 1995) which can hold information in mind for several minutes, suggesting a timescale of WM which is in contrast to other components of WM where information decays over a few seconds. These models have been supported by experimental and neuropsychological case studies (Baddeley, 2000). Despite the evidence that WM is a multi-faceted process involving different brain regions, and that WM is compromised in FM, it is not currently known which dimensions of WM are affected in FM. According to Gil-Ugidos et al. (2021), considering the multiple components of WM and the lack of clarity currently about WM deficits in FM, it is important for research to attend to each of the different dimensions of WM in FM. Specifically for this study, it is not currently known what timescale of WM is typically affected in FM. This is important to understand as evidence-based neuropsychological interventions are based on detailed characterisations of cognitive deficits drawn from the literature (Bilder, 2011) so a more precise conceptualisation of the timescale of WM difficulties in FM is needed to inform appropriate intervention. At present, although there are assessments of WM on a short timescale up to a few seconds used in clinical practice (Lange, 2011) and research into WM deficits in FM (Berryman et

al., 2013), there are none which measure WM on a longer timescale of 10s of seconds to minutes which have been validated for clinical or research purposes. This limits research into the characterisation of WM deficits in FM, meaning that clinical practice is not supported by a robust evidence base. In order to contribute towards addressing this gap in the literature, this study developed and evaluated validity of the first measure of WM on a scale of 10s of seconds to minutes for use in FM populations, which can be used in clinical research to investigate WM deficits in FM comprehensively.

In order to achieve this aim, this study focused on developing and validating a measure for 'top-down' WM. Top-down has been defined as cognitive processes subserved by the prefrontal cortex which enhance and suppress aspects of information processing as required for goal attainment (Gazzaley et al., 2005). It has also been defined as 'cognitive influences and higher-order representations' which have an impact on information processing at earlier, 'bottom-up' cognitive stages (Gilbert & Li, 2013). Pertinently for this study, top-down cognitive processes have been demonstrated to influence how we attend to information in line with previous experiences, whereby we develop expectations and hypotheses to be tested through subsequent experiences (Raviv et al., 2012). Such top-down influences underpin a cognitive process called the contraction bias. The contraction bias occurs when an individual is presented with multiple stimuli with varying magnitudes, and develop a perceptual bias towards the mean of these stimuli, relying on an accumulation of information in WM. They then overestimate the magnitude of relatively smaller magnitude stimuli and underestimate the size of relatively larger magnitude stimuli in the presented sample (Ashourian & Loewenstein, 2011). The measure developed for this study utilised a cognitive bias which is closely linked to this phenomenon, termed the Time Order Effect (TOE), which has been recognised and researched since the 19th Century (Hellstrom, 1979), and involves top-down influences on parametric WM. In common with all cognitive biases, the TOE refers to an automatic, unconscious cognitive process (Hassleton & Nettle, 2006), where repeated

exposure to stimuli with varying magnitudes produces a perceptual bias towards the mean magnitude, resulting from the time-order in which two stimuli with differing magnitudes are presented. While different theories have been proposed for how the TOE is produced, there is most evidence to suggest that it is secondary to a degradation of memory traces over time, which is an implicit cognitive process (Harrison et al., 2017). This effect has been identified across a range of stimulus parameters including length, loudness, duration, brightness, and most relevant to the current study, frequency (Harrison et al., 2017). It has also been identified across sensory modalities (Harrison et al., 2017), and an auditory version of the TOE was used in this study, enabling participants to complete the tasks using their own devices at home. This was essential as the recruitment was conducted during the COVID-19 pandemic which placed restrictions on face-to-face research. The TOE is described in more detail in the Methods section of this report.

This study attempted to validate the measure by administering it across patient and control groups and looking for replicability of findings. Validity has been defined as the accuracy of a measure (Borsboom et al., 2004), and can be subdivided into two types, internal and external validity (Patino & Ferreira, 2018). Internal validity refers to whether the experiment was designed, conducted and analysed in a manner whereby the outcomes can be trusted as a measure of what is intended, meaning the effect of independent variables on dependent variables, not confounding or extraneous variables (Andrade, 2018). One type of internal validity which is particularly relevant for the measure to be assessed in this study is construct validity, where the measure produces outcomes which are consistent with underlying theoretical expectations (Hays & Reeve, 2008), evidencing that the outcomes accurately reflect the constructs they are intended to assess (Borsboom et al., 2004). In this study, the construct to be measured is WM via the auditory modality (AWM) on a timescale of 10s of seconds to minutes. External validity refers to whether the outcomes from research can be generalised outside the research setting. The type of external validity most relevant to this study is population validity, where the findings

can be generalised to multiple relevant groups of people (McDermott, 2011). Both internal and external validity are important considerations for determining the efficacy of a measure, and differentiation between the two types of validity is important for understanding the underlying concept of interest (Fabrigar et al., 2020).

External validity is important to establish because all experiments are context-dependent, and the usefulness and meaning of the outcomes of a study are limited unless generalizable to other scenarios (Aronson et al., 1990). As stated by McDermott (2011), external validity as established by replication across population, context and time should be sought prolifically, because if there is a genuine causal relationship between the independent and dependent variables in a study, it should emerge over these different dimensions. The primary focus of the current study was therefore on external validity as a key initial step towards producing a useful measure for clinical research and practice. Specifically, it focused on population validity, to help determine whether the measure has validity to test AWM at a scale of 10s of seconds to minutes across a wider and more meaningful scale. To achieve this, the measure was administered with identical methodology to different samples with differing clinical status, and the patterns of outcomes were examined for continuity and similarity. Assessing population validity requires administering the measure to different groups who would be expected to be representative of the construct under investigation (McDermott, 2011). As the measure was to be validated for use in research into FM, and the TOE has been observed in healthy individuals previously (Harrison et al., 2017), the groups in this study were a Healthy Control (HC) group and a FM group. This method was chosen as assessing external validity via reproducibility across different populations is a more effective means of establishing external validity than other methods such as increasing the breadth or size of clinical samples within a single study (McDermott, 2011). Patterns of outcomes were expected to be similar for each group, notwithstanding some variation as WM capacity is already established as a general area of difficulty for people with FM. More specifically, we expected the within-subject effects

for each group to have the same pattern and be statistically significant, although the effects may be weaker in the FM group because of WM deficits impacting on performance. We considered population validity to be evidenced if a similar pattern occurred for each group, even if the pattern occurred to different extents between groups. As described in more detail in the Methods section, the measure exposed participants in a repeated measures design to differing compositions of frequencies of auditory stimuli, which invoked TOEs with skewed patterns of accuracy and reaction time. More specifically, in relation to the standard TOE paradigm, in which there is a 50:50 ratio of stimulus trials, ratios of 20:80 and 80:20 of the same stimuli were expected to produce varying magnitudes of the TOE across different trials. For trials which occurred with a 20% probability in each of these blocks, an increased TOE magnitude was expected while a reduced TOE was expected for trials which occur with an 80% probability in these blocks. In terms of establishing external population validity, this pattern of results would be expected consistently across the FM and HC groups, as well as both groups combined.

While the primary aim of the study was to establish whether the novel measure had evidence of population validity, these outcomes would additionally provide evidence towards construct validity. Construct representation, which is an aspect of construct validity theory, refers to how well experimentally manipulated outcomes are an indicator of the underlying construct (Borsboom et al., 2004). As such, the aim of construct representation is to test whether the theorised underlying construct produces the theorised pattern of dependent variable outcomes, with construct representation being evidenced by confirmation of the predicted outcomes (Strauss & Smith, 2009). If the anticipated outcomes previously described were elicited by the novel measure of AWM developed in this study, it could therefore be concluded that there was evidence of construct validity.

In addition to the population and construct validity assessed as primary foci for this study, preliminary evidence was sought concerning the criterion validity of the novel measure. Specifically, a subtype of criterion validity called concurrent validity was examined, which refers to whether a new measure's

accuracy can be corroborated by comparison with existing 'gold-standard' measures of the same or similar constructs (Davis et al., 2017). In order to examine concurrent validity, this study compared the outcomes of the novel measure of longer-timescale AWM with the Weschler Adult Intelligence Scale Working Memory Index (WAIS-IV WMI) and n-back task, which are established clinical and experimental measures of general WM respectively (Holdnack, 2019; Gajewski et al., 2018). As previously stated, there is currently no measure of WM at a timescale of 10s of seconds to minutes, hence the need for this new measure, and the WAIS-WMI and n-back measure WM at a timescale of a few seconds. However, WM is not a unitary construct and is hypothesised to include multiple components incorporating different dimensions including timescale, and these components interact. Most pertinently for this study, early-stage WM processing is hypothesised to be inter-dependent with longer-timescale WM (Ricker et al., 2010). As such, although the WAIS-IV WMI and n-back measure different timescales of WM to the novel AWM task, it was expected that there would be evidence of a partial correlation between these measures. This finding would provide preliminary evidence of concurrent validity of the novel measure.

Considering the potential implications of this study for clinical assessment, preliminary evidence of a correlation between the n-back task, and the WAIS-IV WMI, was also analysed. As both measures assess general WM on a similar timescale, it was expected that they would be positively correlated. While the n-back task used was based on standard versions of the test which already have demonstrated validity, the version in this study was adapted for an online administration. Correlating this version with the WAIS-IV WMI would contribute evidence towards the criterion validity of this n-back, specifically concurrent validity. Both measures are described in the Procedure section later in this document. Finally, the study provided pilot data concerning differences between people with a diagnosis of FM and HCs regarding top-down influences on AWM. In line with the evidence outlined so far indicating that

WM in general is an area of difficulty in FM, it was expected that top-down influences in the FM group are impaired relative to the HC group.

We hypothesised that:

1. The TOE will be measurable with respect to parametric WM in both clinical and control groups. More specifically, performance in terms of accuracy and reaction times on the AWM task will be influenced by expectations learnt from previous trials where auditory information is presented over a longer period. In terms of establishing evidence of construct validity of the measure, we expect that in trials which occur with 20% probability, there will be a consistent and statistically significant pattern of an increased TOE relative to the baseline 50:50 ratio condition. There will also be a statistically significant pattern of decreased TOE relative to the baseline condition for trials which occur with an 80% probability. Demonstrating these patterns consistently across both groups individually as well as when both groups combined would provide evidence for population validity.
2. We additionally anticipate that, when comparing participants with FM to HCs, there will be preliminary evidence of deficits in top-down influences on AWM and general WM capacity. However, this difference will not detract significantly from the expected pattern of outcomes across the two groups as stated in hypothesis one.
3. We anticipate that there will be further preliminary evidence that individual differences in general WM ability will partially explain individual differences in AWM capacity. In particular, it is expected that the outcomes of longer-term AWM as described in the Methods section of this study will correlate with a small to medium effect size with the established measures of general WM, the WAIS-IV WMI and n-back. This would provide preliminary evidence towards concurrent validity of the novel measure.

4. We finally expect to find preliminary evidence of a positive correlation between the n-back task and the WAIS-IV WMI. If this is found, it would contribute evidence towards the concurrent validity of the n-back task.

Materials and Methods

1. Experimental Design

This quantitative research study used a between-subjects observational (case-control) design with a nested within-subjects experimental design.

2. Participants

2.1. Participant identification and initial contact

Participants with a diagnosis of FM living in the community were recruited, who had been involved in previous studies with the University of Liverpool and expressed an interest to be retained on a register for future research. A further healthy participant group which was age-matched to the clinical group was recruited from University of Liverpool staff and student populations, as well as the wider population. Participants were provided with an information sheet before deciding whether to take part in the study. This sheet explained that the study was investigating the role of WM in FM, and would contribute to a better understanding of these widely reported difficulties to inform clinical practice. The information sheet also explained that the study would involve a computer experiment, conducted in their own home, which would involve listening to and making judgements about auditory and visual stimuli. Participants were given further opportunities to ask follow-up questions directly to the researchers by phone or video call if they wished.

Patient participants were identified through a database of participants held by the Chief investigator. These patients had previously taken part in research at the University of Liverpool. This included both patients with FM and healthy individuals. Potential participants were contacted directly by the study team using retained contact details gathered from previous studies (See Appendix F and Appendix G). They were only contacted if they had previously expressed an interest in taking part in further research and provided written consent to be contacted for this purpose.

2.2. Eligibility criteria

The inclusion criteria were: (i) clinician-confirmed diagnosis of FM (clinical group) or no history of CP (healthy group); (ii) being over 18 years and under 65 years of age. The exclusion criteria were: (i) comorbid neurological impairments such as brain injuries or neurodevelopmental disorders; (ii) failure of the test of effort (as described in the Procedure section later in this report). Although normal or corrected-to-normal hearing was not formally an inclusion criterion, all participants were asked about their hearing and none reported difficulties, or struggled in conversation before undertaking the test.

2.3. Eligibility screening

Potential participants who expressed an interest in taking part in the study were contacted by C.K., Trainee Clinical Psychologist, who is one of the researchers, to determine whether they met inclusion or exclusion criteria. This study was conducted remotely, with participants requiring a computer and reliable internet connection in their own home. This was discussed during screening and eligibility phone calls.

2.4. Ethics and patient safety and expert by experience consultation

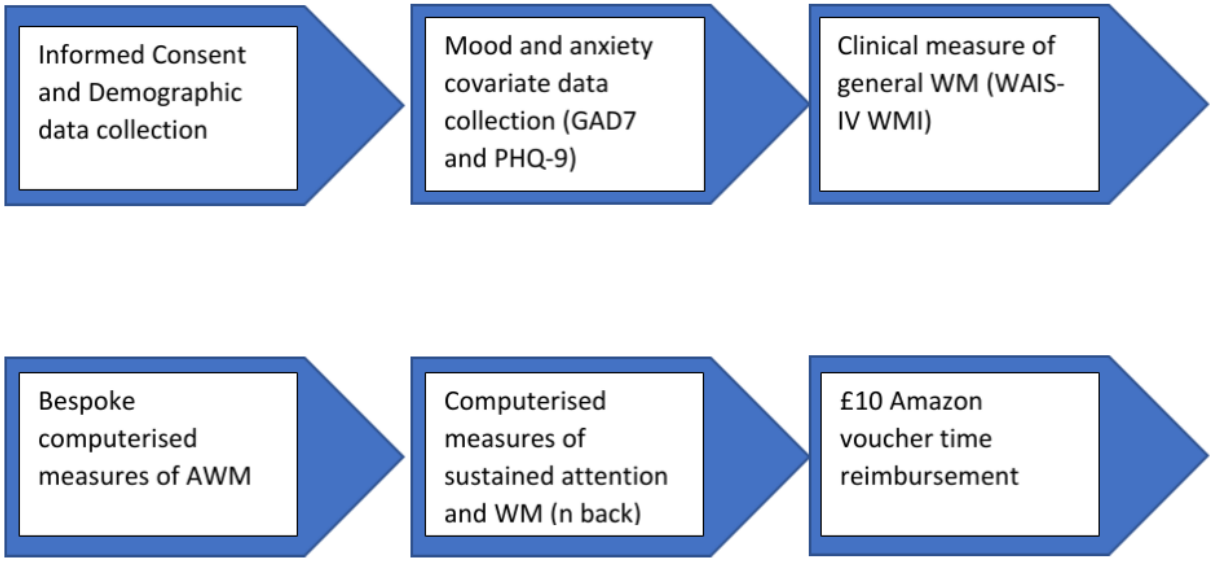
This study was granted ethical approval by the University of Liverpool Sponsor on 21st August 2020, the Higher Research Authority (HRA) and Health and Care Research Wales (HCRW) on 15th January 2021, and the NHS Research REC on 15th January 2021 (see Appendices H-J for Statements of Approval). All participants were given information about the study and provided informed consent before taking part.

An expert by experience provided detailed feedback on the project proposal, and amendments were made in accordance with these. While the exact nature of the expert by experience's health condition was not specified and the feedback was anonymous, they did state that they experienced WM difficulties. It was specified in the project proposal submitted for consideration by the Liverpool Doctorate in Clinical Psychology department and the NHS Research Ethics Committee (REC) that further consultation would be undertaken with the Self Motivation in Lasting Endorphins (SMILE) Liverpool charity for people living with CP, by attending their meetings and forming focus groups. However, due to restrictions on non-essential gathering in the context of COVID-19, SMILE did not meet over the subsequent time period and it has unfortunately not been possible to consult further with them.

3. Procedures

3.1. Experimental procedure

Figure 1. Procedure Overview



Following informed consent gathering, participants were asked to return a questionnaire with key demographic and medication information (See Appendix K). Once this information was returned, a video call was arranged between the researcher and each participant. During the call, the Generalised Anxiety Disorder Assessment (GAD7; Appendix L) and Patient Health Questionnaire 9 (PHQ-9; Appendix M) were then administered as standard measures of anxiety and depression respectively. The Weschler Adult Intelligence Scale IV, WM Index (WAIS-IV WMI), a widely used clinical measure of general WM was then administered. The WAIS-IV WMI is fully verbal and was administered by video call between the participant and a researcher. The task related to AWM was then administered to all participants using bespoke computer programmes generated with Gorilla online software, which produced patterns of auditory tones. Following this, all participants completed a computerised n-back task using Gorilla online software. All of the computer experiments had regular break points, and participants were encouraged to use these to manage

any fatigue or pain they were experiencing. This was reported to be particularly helpful for those participants who experienced finger or hand pain.

3.2. Measures

The following measures were used in this study:

3.2.1 The Generalised Anxiety Disorder Assessment (GAD7) and Patient Health Questionnaire 9 (PHQ-9) were administered as measures of anxiety and depression respectively. The GAD7 has excellent internal consistency (Cronbach alpha=0.92), good test-retest reliability (intraclass correlation=0.83), as well as good construct, factorial, criterion and procedural validity (Spitzer et al., 2006). The PHQ-9 has high internal consistency (Cronbach alpha between 0.86 and 0.89), as well as good criteria validity (APA, 2020), good construct validity and acceptable test-retest reliability (test-retest correlation coefficient=0.737; Sun et al., 2020). The GAD7 has seven items, each of which has a score of 0-3, where increased scores indicate higher anxiety. As such the GAD7 has a range of 0-21, and a score of over 10 indicates Generalised Anxiety Disorder (Spitzer et al, 2006). The PHQ-9 has 9 items, each of which has a score of 0-3, where higher scores indicate greater depression. As such the PHQ-9 has a range of 0-27, and a score of over 10 has a sensitivity of 88% and specificity of 88% for major depression (Kroenke et al., 2001). Depression and anxiety are frequently reported to negatively impact on cognitive domains including attention, memory and processing speed (Perini et al., 2019), meaning that any variance in the cognitive processes measured in this study may be attributable to affective factors in addition to CP. As such they were considered as covariates in this study.

3.2.2 The Weschler Adult Intelligence Scale IV, WM Index (WAIS-IV WMI). The WAIS-IV WMI has been normed on a stratified sample of 2200 healthy individuals, aged from 16 to 90, with further stratification in respect to gender, race, ethnicity, educational level and geographical region. It

has high face validity and good reliability (Chronbach's alpha of 0.94: Silva, 2008). The WAIS-IV WMI has not been tested with FM patients, but would be valuable for clinical application of any findings of WM deficits in this population, particularly for assessment in clinical settings. In addition, the WAIS-IV WMI contains an embedded test of effort termed the Reliable Digit Span measure (Greiffenstein, Baker & Gola, 1994). Tests of effort are commonplace in neuropsychological test batteries, as reduced effort is well documented to impact on task performance (e.g. Green, 2007), and as such needed to be accounted for in this study. The WAIS-IV WMI has two subtests; Digit Span and Arithmetic. In Digit Span, participants listened to strings of numbers, which increased in length, and were asked to repeat the strings. In trial one, the string was repeated in the same order as the experimenter states it, while in trial two, they were asked to repeat it in reverse order, and in trial three they were asked to repeat it but ordering the numbers from lowest to highest. In the Arithmetic subtest, participants listened to an arithmetic problem which required holding verbal information in mind. In both subtests, the raw score calculated was total correct responses, which were then converted to scaled scores based on normative data. The two scaled scores were combined to produce a composite score, (Mean=100, Standard Deviation=15, which was the outcome measure used for analysis.

3.2.3 A computerised visual version of the n-back task. The n-back task is a widely used measure of general WM with good face validity (Gajewski et al., 2018), and moderate reliability (Chronbach's alpha of 0.55: Shamosh et al., 2008). Previous research has found the n-back task to be the most sensitive measure of cognitive difficulties in FM (Seo et al., 2012). It requires the participant to identify whether a visual stimulus matches another presented several steps previously, or a target visual stimulus. As such, it engages several key WM processes, including sustained attention, on-line monitoring, updating and manipulation of information. A visual letter version which has been used and validated in CP populations (Moore at al., 2019) was

used. This stage included two versions of the n-back task: (i) a 0-back task, where participants had to state whether a briefly presented letter matched a predefined letter (X) or not, which was used to give an indication of sustained attention capacity; and (ii) a 2-back task, where participants were required to respond whether a presented letter was the same as one presented 2 trials previously. As such, this version loaded more heavily on WM capacity specifically than the 0-back task. Each version had 100 trials, stimulus presentation was 200ms, the response window was 3200ms (+/-50ms), with 30 target stimuli out of the 100 trials. Participants responded by using their keyboard to indicate on each trial whether the target stimulus was presented or not.

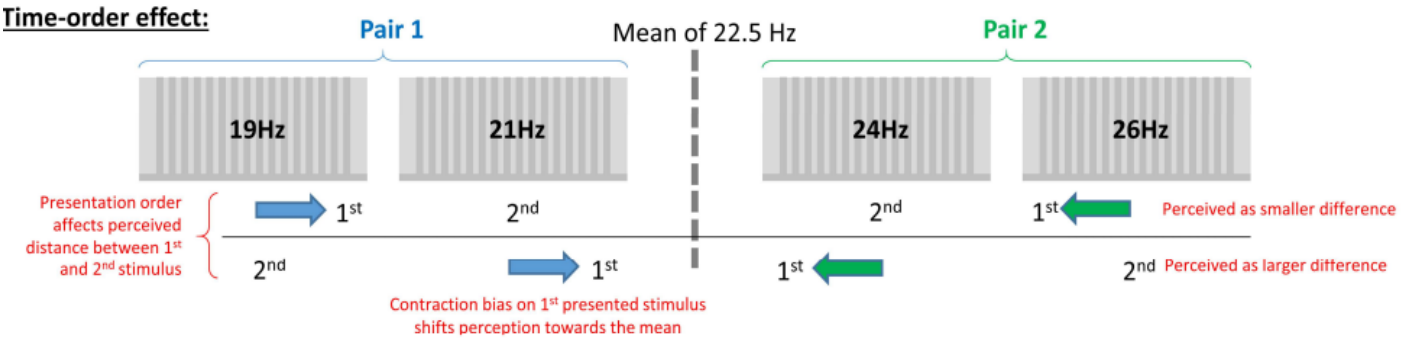
3.2.4 A bespoke computerised programme created for this study measuring AWM capacity (described below).

3.2.4.1 Development of novel AWM Measure

If an individual is repeatedly exposed to four stimulus trains with different frequencies, for example 19Hz, 21Hz, 24Hz and 26Hz, over time a perceptual bias towards the mean of those frequencies will occur (Hellstrom & Rammsayer, 2015), which is 22.5Hz in this example. This manifests cognitively as an expectation that the four distinct stimulus trains will be closer to 22.5Hz than they actually are in further trials (see Figure 2). Behaviourally, if the individual is then presented with the four stimulus trains in pairs (19Hz/21Hz and 24Hz/26Hz), and asked to discriminate between the frequencies, the responses become skewed in terms of accuracy depending on which frequency is presented first. In this example, if the lower frequency pair is presented with the order 19Hz – 21Hz, a lower accuracy would be expected than if it was presented in the order 21Hz – 19Hz. Conversely, if the higher frequency pair is presented in the order 24Hz – 26Hz, a higher accuracy would be expected than if the same pair was presented in the order 26Hz – 24Hz. This reflects the overall bias towards the mean of 22.5Hz in that when the stimulus train frequency pair which is closest to the mean is presented first, there is a greater perception of difference to the second

stimulus train, which is further from the mean, and consequently discrimination accuracy is higher. This effect is called the TOE in the literature and, using the frequencies described above, has been previously demonstrated in pilot studies using tactile stimuli (See Appendix N). While the exact mechanisms underlying the TOE remain controversial, it is an established phenomenon which highlights different levels of WM and their interaction. In tasks investigating the TOE, basic WM processes are required to discriminate between two sequentially presented stimulus trains in a short period of time. Concurrently, top-down WM processes are required to hold information about multiple stimulus trains over a longer time period in mind, and form probabilistic expectations for current and future stimulus trains based on those previous trains. The top-down processes furthermore use the information held in mind over a longer time period to influence the perception of the stimulus trains in trials where more rudimentary WM processes are primarily involved.

Figure 2. Top-down (time-order) effects on AWM



The measure developed for this study manipulates the occurrence of stimulus frequency pairs as outlined above in order to assess AWM capacity across different timescales. The measure makes use of varying blocks with either 50:50 or 80:20 ratio compositions of stimulus pairs centred around 20Hz or around 25Hz. This is described in more detail in the Bespoke AWM Task Structure section of this report. The purpose of altering the block compositions in this way was to vary the unconscious probabilistic

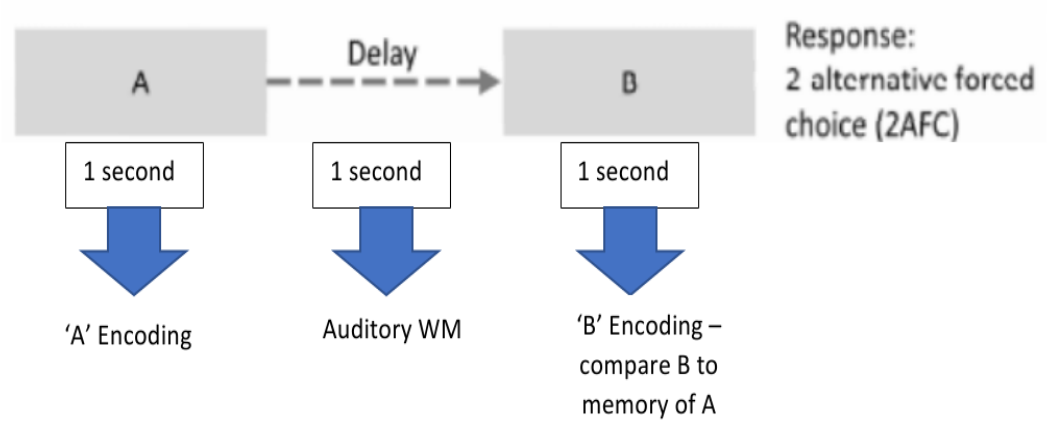
expectations elicited according to the TOE, which rely on AWM at different points. For example, in a block with a 50:50 composition of stimulus train pairs, the standard TOE was expected, as the mean frequency expectation would be 22.5Hz. However, in the 80:20 ratio blocks, where 80% of stimulus pairs centre on 25Hz, and 20% centre on 20Hz, the mean frequency (24Hz) is closer to the 25Hz pair. In these blocks, it was expected that the TOE would be more pronounced for the pair of stimulus trains centred on 20Hz, as they are now further perceptually from the block mean, and that the TOE for the 25Hz centred pair would be diminished, as the mean is now close to 25Hz. Conversely, in 20:80 ratio blocks, with 20% of stimulus pairs centred on 25Hz and 80% of stimulus pairs centred on 20Hz, the mean frequency would be 21Hz, producing a more pronounced TOE in the 25Hz pairs and a diminished TOE in the 20Hz pairs. Pilot data conducted using these frequency manipulations using auditory stimulus trains and healthy participants for this study suggested that this phenomenon was elicited (See Appendix O). As each block lasts several minutes, longer-term AWM lasting in the scale of 10s of seconds to minutes is required to see this effect. The retention of stimuli in AWM over the course of the full 80:20 or 20:80 block is necessary to produce the different mean frequency and probabilistic expectations which skew the TOE responses as described. In contrast, the TOE in the 50:50 ratio blocks could be elicited after relatively few trials, requiring AWM at a lower level and over a shorter timescale.

These variations in blocks were anticipated to produce differences in response accuracy, and also reaction times, which were the quantitative outcomes. More specifically, the differences in accuracy and reaction times between the 50:50, 80:20 and 20:80 blocks were anticipated to provide an indication of whether an individual had difficulties with AWM across a range of timescales from a few seconds to minutes.

3.2.4.2. Current Task Structure

Within the 50:50 blocks of the computer-generated experimental measures described above, existing methods were used that produce robust results in HCs (Romo & Salinas, 2003), but were being used with people with FM for the first time in this study. The addition of 80:20 and 20:80 ratio blocks was a modification to existing procedures which was being introduced for the first time, and was the one to be validated in this study. In order for participants to perform optimally on this task by maintaining accuracy and speed in their judgments, they would make use of prior information, maintained over more trials than the 50:50 blocks and therefore requiring maintenance in WM over a longer timeframe. Deficits in participants' utilization of this prior information would indicate a deficit in longer-term, top-down WM. The within-trial structure for all trials across the bespoke AWM task is visually represented in Figure 3. Before undertaking these tasks, participants were given options for which computer keys felt most comfortable to use and completed a calibration task using the same trial format as the one described previously and outlined in Figure 3. In this calibration task, participants were exposed to pairs of auditory stimulus trains which had a mean frequency of either 20Hz or 25Hz, as with the experimental trials described later. However, in the calibration trials the difference between the frequencies in each stimulus pair became progressively smaller block by block, and therefore progressively harder to discriminate. Across four blocks of 20 trials which all participants completed with this format, the appropriate level of difficulty was selected for each participant, according to a 75% accuracy cut-off, where participants would progress to an experiment with the smallest – and therefore hardest – frequency differences where they scored over 75% during calibration. This recognised a finding from early pilot runs of the experiment where different pilot participants found different frequencies more or less challenging and ensured that participants did not find the task too easy or too difficult, which could have undermined the results.

Figure 3. Within-trial structure across all trials of the AWM task



Following the calibration task, participants progressed to an experiment where they were exposed to 120 trials of stimulus pairs, divided into six blocks of 20 trials each. Each trial lasted between three and eight seconds depending on participant response time, with each block therefore lasting between 60 and 160 seconds. Each stimulus pair had a mean frequency of either 20Hz (low frequency pair) or 25Hz (high frequency pair). The blocks were comprised of ratios of 50:50, 80:20 and 20:80 with reference to high frequency pairs: low frequency pairs. There were two blocks of each ratio, producing three conditions of the 50:50, 80:20 and 20:80 ratios, with 40 trials in total per condition. This was based on literature which indicates that over 30 trials are required to produce the TOE, and pilot data which indicated that 40 trials split over two blocks per condition was as effective at producing the TOE as 40 trials per block (see Appendix O). The experiment was counterbalanced to control for order effects and participants were accordingly randomised into either an ABBA or BAAB design (see Table 1). The 50:50 ratio blocks and five-minute break were placed in anticipation of the potential skewing of response accuracy and reaction times in the 20:80 and 80:20 blocks. Therefore the 50:50 blocks and break were positioned to washout the cognitive biases introduced by the tasks before proceeding to the next block.

Table 1. Counterbalancing formats

ABBA Format (High:Low frequency pairs)	BAAB Format (High:Low frequency pairs)
1. 80:20	1. 20:80
2. 50:50	2. 50:50
3. 20:80	3. 80:20
4. 5 minute break	4. 5 minute break
5. 20:80	5. 80:20
6. 50:50	6. 50:50
7. 80:20	7. 20:80

This task was designed to (i) measure implicit simple integration of temporal auditory information in AWM. All of the trials of sequential trains of auditory stimuli were presented at a speed that was too rapid for each individual stimulus in the train to be fully processed consciously, and the participant was asked to judge which train as a whole was of faster frequency (see Figure 4). This depended on parametric AWM processes.

Figure 4. Measure of implicit AWM

The task also measured (ii) the impact of prior learning and top-down cognitive processes on AWM across all trials. As participants were presented with sequential trains of auditory stimuli which differed in terms of frequency within blocks and across trials, the TOE was elicited, which depends on longer term AWM on a scale of seconds to 10s of seconds to amaintain auditory information in short-term memory stores while

adjusting responses in line with the representation of average frequency across previous trials. This means they needed to make use of their learning of the distribution of frequencies and likelihood of changes in frequencies when making judgements about the current stimulus trains. This prior learning is thought to engage high-order WM processes due to the need to retain information over longer timescales than the simple integration level of WM required for individual trial discrimination.

Finally (iii), participants' variation in performance across different blocks of stimulus trains with varying compositions of stimulus frequencies provided a further indication of how top-down processes impact on implicit AWM. Differences between blocks where ratios of 50:50, 80:20 and 20:80 were presented provided more detailed information concerning the presence and manifestation of the TOE, considering different ratios of stimulus frequencies would be expected to produce different outcomes. This would elucidate the impact of the TOE on responding across different timeframes (10s of seconds to minutes) and with different probabilistic expectations.

3.2.4.2 Quantitative Outcomes

3.2.4.2.1 AWM and WM capacity

In both groups and across experimental conditions, accuracy and reaction time were measured when responding to a forced choice concerning which of two trains of auditory stimuli was faster. An overall percentage accuracy score was computed for each participant based on correct responses divided by the total number of attempted trials, multiplied by 100. A similar calculation was conducted to calculate percentage accuracy for the 0-back and 2-back tasks separately. Participant responses were made by keyboard press on their own home computer. Accuracy and reaction time were recorded by the Gorilla online software which was being used to deliver the experimental trials. This resulted in four Dependent Variables (DV) in total, which are described in detail in the Data Processing section.

3.2.4.2.2 General WM capacity and effort

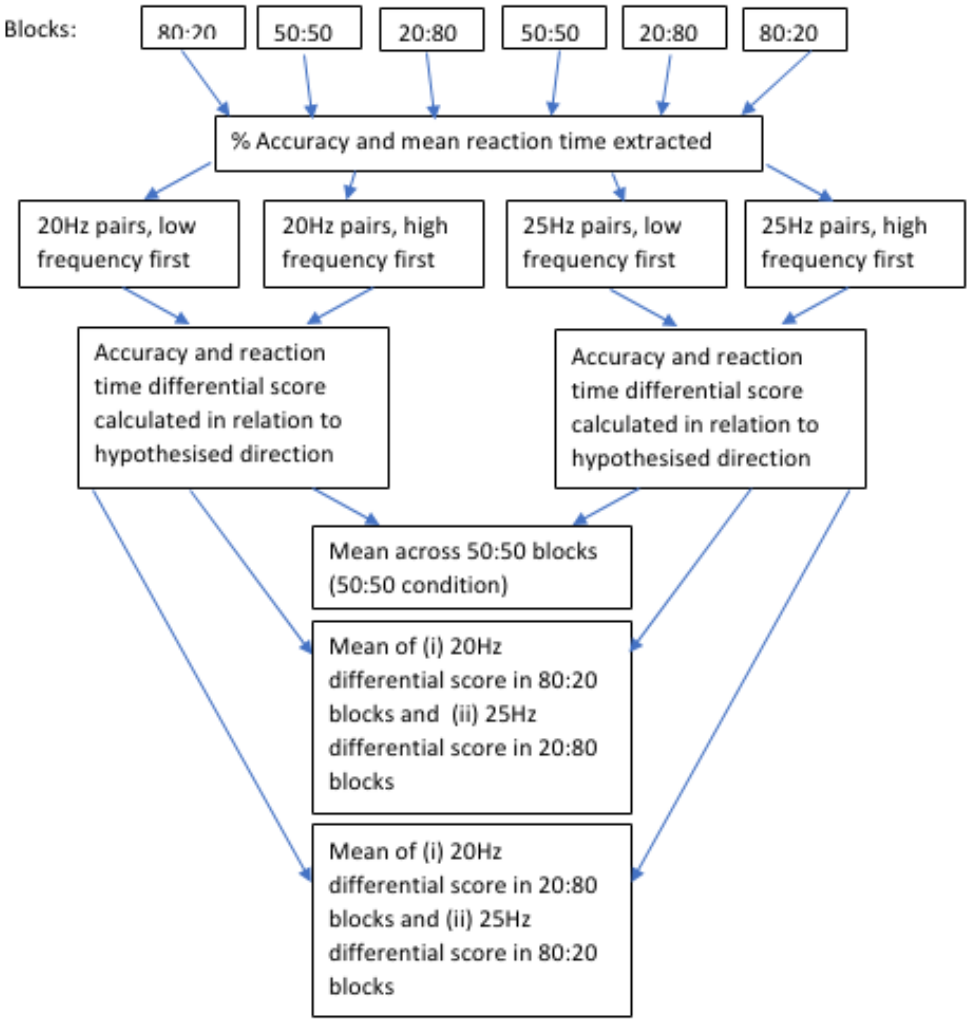
The general WM DV derived from the n-back task was accuracy, captured by the Gorilla online software which delivers the task. The general WM DV derived from the WAIS-IV WMI was normed percentile. Effort measure captured by the Reliable Digit Span was determined as below or above clinical cut-off scores.

3.2.4.2.3 Mental Health covariate data

The GAD7 and PHQ-9 provided clinical measures of anxiety and depression respectively. Both scales provided ordinal outcome data.

2.2.4.2.4 Data processing

Figure 5. Data Processing Visualisation



The correct responses and reaction time for each AWM task trial were extracted and divided into each of the three block types. Percentage accuracy and mean reaction time were calculated for each block, according to which stimulus train pair was being presented, and which frequency was being presented first within the stimulus train pair. This resulted in four accuracy (%) and four reaction time (milliseconds converted to logarithm) scores for each block: stimulus train pairs centred on 20Hz when the slower frequency was presented first; stimulus train pairs centred on 20Hz when the faster frequency was

presented first; stimulus train pairs centred on 25Hz when the slower frequency was presented first; stimulus train pairs centred on 25Hz when the faster frequency was presented first.

Subsequently, the percentage accuracy and reaction time *difference* for each stimulus train pair was calculated across each block and for each participant. This resulted in two accuracy differential scores and two reaction time differential scores per block: difference in percentage accuracy and reaction time between trials where the slower frequency was presented first and where the faster frequency was presented first for trials with the stimulus train pair centred on 20Hz; and then again for the stimulus train pair centred on 25Hz. Trials where the first presented frequency was closest to the block mean will be termed 'near' and trials where the first presented frequency was furthest from the block mean will be termed 'far' from hereon. The direction of the difference was also calculated with reference to the expected, hypothesised direction. For instance, according to the literature on the TOE and the pilot data gathered, during the 50:50 condition trials where the stimulus train pairs centred on 20Hz were presented, it was expected that accuracy would be lower for trials where the lower frequency was presented first, while the reaction time would be higher. If this was the outcome for a participant, their score was recorded as a positive number. If an individual's outcome was the opposite to expectations - for instance in the trials mentioned above, the accuracy for trials where the slower frequency was presented first was higher than for trials where the faster frequency was presented first or the reaction time was lower - the outcome was recorded as a negative number.

Following this, the *means across blocks* of these relative differential percentage accuracy and reaction time scores were calculated for each participant according to the directions of change which were of interest to this study. More specifically, the following means were calculated:

a. The means across the average relative differential percentage accuracy scores (and separately, the differential reaction time scores) in the two 50:50 blocks. This produced the condition which will be termed the 50:50 condition from hereon.
b. The means across (i) the differential percentage accuracy score for the trials centred on 20Hz in the 80:20 blocks and (ii) the differential percentage accuracy score for the trials centred on 25Hz in the 20:80 blocks. This was repeated for reaction times. This produced the condition which will be termed the 20% probability condition from hereon.
c. The means across (i) the differential percentage accuracy score for the trials centred on 25Hz in the 80:20 blocks and (ii) the differential percentage accuracy score for the trials centred on 20Hz in the 20:80 blocks. This was repeated for reaction times. This produced the condition which will be termed the 80% probability condition from hereon.

In summary of the above, the DVs produced in this experiment were (i) differential percentage accuracy scores and (ii) differential reaction time scores. The independent Variable (IV) was the stimulus train pair probability, which had three levels (i) 50:50 condition, (ii) 80% condition and (iii) 20% condition.

4. Statistical Considerations

4.1. Sample size

For analysis relating to the within-subjects factor with three levels, stimulus train pair probability: With regards to the repeated measures ANOVA required for validating the measure of longer-term, top-down AWM processes (hypothesis 1), a sample size of 15 was required, to detect a large effect (f=0.8) at 80% power and an alpha of 0.5 with a 2-tailed hypothesis. Hellstrom & Rammsayer (2015) found that the TOE in within-subjects analyses across two distinct experiments was produced with large effect sizes, warranting the large effect size. This study aimed to recruit 17 participants with FM to ensure robust findings. In addition, the experiment was repeated in a control group of 17 HCs to test reproducibility. An additional

within-subjects analysis was also conducted by pooling all participants across patient and HC groups, which was suitable for detecting a medium effect size (f=0.50) at 80% with an alpha of 0.5 and a 2-tailed hypothesis. This analysis compared conditions where participants responded to stimuli designed to produce a standard TOE to conditions where stimuli were presented which were designed to produce increased or decreased magnitudes of the TOE (described in more detail in the Materials section).

For analysis relating to the between-subjects factor with two levels, which was experimental group (FM and HC): With regards to the between-groups analysis of differences in WM and AWM capacity when comparing patients and HCs (hypothesis 2), as well as between groups differences in top-down AWM processing (hypothesis 3) and correlations between outcomes (hypothesis 4) statistical significance was not sought, as its purpose was to provide pilot data and indications of effect size. As such, a control group of 17 healthy participants was recruited to match the experimental group, making a total of 34 participants.

4.2. Data processing

The Gorilla online software produced a spreadsheet for each participant, for the n-back task and the AWM task separately. Overall percentage accuracy on the AWM task was calculated initially and any participants who scored below 50% were excluded from the analysis, as their scores were at or below chance. This resulted in data from two HCs and one participant with FM being excluded from the analysis. It was also noted that one participant with FM's age had been improperly recorded at the screening stage, and they were in fact outside the inclusion age range. This participant's data was excluded, meaning 30 participants were included in the final analysis, 15 in each of the FM and HC groups.

Results

1. Demographic, mood and fatigue data gathered at the start of the study is presented in Table 2.

Table 2. Demographic mood and fatigue information

	FM Group (n = 15)	HC Group (n = 15)	Combined FM and HC (n=30)
Mean age in years (SD; range)	53.9 (7.6; 42-64)	53.5 (8.7; 37-63)	53.7 (7.9; 37-64)
Gender (Female:Male)	13:2	12:3	25:5
Mean PHQ-9 score (SD)	12.9 (4.7)	2.7 (3.1)	7.8 (6.5)
Mean GAD7 score (SD)	9.3 (4.8)	1.7 (1.8)	5.5 (5.3)
Fatigue rating out of 10, where 10 is highest (SD)	6.8 (1.5)	2.6 (2.6)	4.7 (2.9)

Participants in the FM group generally had higher depression and anxiety scores, as well as reporting greater fatigue at the time of testing. An independent samples t-test indicated that on the GAD7 the 15 participants in the FM group (M=9.3, SD=4.8) reported significantly higher scores compared to the 15 participants in the HC group (M=1.7, SD=1.8), $t(28)=5.755$, $p<0.001$. A further independent samples t-test indicated that on the PHQ9 the FM group (M=12.9, SD=4.7) reported significantly higher scores compared to the HC group (M=2.7, SD=3.1), $t(28)=7.038$, $p<0.001$.

2. Data analysis:

2.1. Initial one sample t-tests were computed for the accuracy and reaction time differential scores in the 50:50 condition to determine whether they were significantly different to the null hypothesis that there is no difference between the 50:50 condition and zero. This was conducted to check the validity of the task in that the standard TOE as described in the literature was produced by the task before moving to specific hypotheses for this study. The results of these t-tests are summarised in Table 3.

Table 3. Independent samples t-tests comparing accuracy and reaction time differential scores in the 50:50 condition compared to zero (null hypothesis)

	FM Group (n = 15)	**HC Group (n = 15)**	**Combined FM and HC (n = 30)**
Accuracy differential score	M = 13.9, SD = 23.1, t(14) = 2.343, p = 0.041	M = 11.9, SD = 15.5, t(14) = 2.954, p<0.01	Mean = 12.9, SD = 19.4, t(29) = 3.652, p<0.001
Reaction time differential score	M = -0.04123, SD = -0.03178, t(14) = -5.024, p<0.001	M = -0.02571, SD = -0.3188, t(14) = -3.124, p<0.01	M = -0.03218, SD = -0.03347, t(29) = -5.683, p<0.001

Across both groups combined, the accuracy and response time differential scores were in line with the hypothesised directions. These findings were consistent across FM and HC groups separately. These findings indicate that the experiment was successful in producing the baseline standard TOE across FM and HC groups, using measures of accuracy and reaction time.

2.2. To test hypothesis one, regarding within-subjects top-down cognitive process effects on AWM: Mixed and repeated measures ANOVAs to analyse within-subjects differences in response accuracy and reaction times, with group included as a between-groups factor in the analysis where the clinical and control groups are pooled, between the 50:50, 20% probability and 80% probability conditions.

2.2.1. Combined Group ANOVA

Initial tests of assumptions indicated that accuracy and reaction times data in all conditions were normally distributed and met assumptions of sphericity.

2.2.1.1. Accuracy outcomes

A mixed ANOVA was conducted where the within-subjects factor was experimental condition and the between-subjects factor was experimental group. There was no significant main effect of group, F (1, 28) =0.17, p=0.683. The within-subjects results showed that the accuracy differential score differed significantly across the conditions with varied probability of stimulus pair occurrence [F(2, 56)=3.115, p=0.05] with a medium to large effect size ($\eta^{2=}$0.101). Post hoc tests (see Table 4) using the Bonferroni correction revealed that the accuracy differential score reduced significantly by 8.92 points in the 80% probability condition compared to the 50:50 condition, with a medium effect size. There was non-significant reduction by 1.67 points from the 50:50 condition to the 20% probability condition with a negligible effect size. There was no significant interaction between group and variation in TOE effects across conditions F(2,56)=0.06, p=0.941 with a negligible effect size (η^2 =0.002). This effect tells us that varying the probability of stimulus pairs had an effect on the magnitude of the TOE regardless of group.

Table 4. Pairwise comparisons for accuracy differential scores between conditions for FM and HC groups combined (n = 30)

Condition (% Accuracy; SD)	**Compared With Condition (% Accuracy; SD)**	**Outcomes (p value; effect size; Confidence Interval)**
50:50 (12.76; 19.61)	20% Probability (11.09; 25.53)	p = 1; d=0.07; 95 CI [-0.64 – 0.79]
	80% Probability (3.82; 13.18)	p = 0.017 d=0.53; 95 CI [-0.19 – 01.26]

2.2.1.2. Reaction time Outcomes

A mixed ANOVA was conducted where the within-subjects factor was experimental condition and the between-subjects factor was experimental group. There was no significant main effect of group F(1,28)=2.22, p=0.147. The within-subjects results showed that the reaction time differential score did not differ significantly across the conditions with varied probability of stimulus pair occurrence [F(2, 58)=2.846,

p=0.073] and there was a medium effect size ($\eta^{2=}$0.089). While non-significant statistically (see Table 5), the reaction time differential score increased by 0.014 points in the 80% probability condition compared to the 50:50 condition, with a medium effect size. The reaction time differential score decreased by 0.006 points in the 20% probability condition compared to the 50:50 condition, with a small effect size. Overall this pattern of results was in the expected direction although not statistically significant. There was no significant interaction between group and variation in TOE effects across conditions [F(2,56)=0.019, p=0.981) with a negligible effect size ($\eta^{2=}$0.001).

Table 5. Pairwise comparisons for reaction time differential scores between conditions for FM and HC groups combined (n = 30)

Condition (Mean reaction time; SD)	**Compared With Condition (Mean reaction time; SD)**	**Outcomes (p value; effect size; Confidence Interval)**
50:50 (-0.0335; 0.0323)	20% Probability (-0.0399; 0.0526)	p = 1; d=0.15; 95 CI [-0.3601 – 0.6534]
	80% Probability (-0.0198; 0.0347)	p = 0.222 d=0.41; 95 CI [-1.132 – 0.3151

2.2.2. FM Group ANOVA

Initial tests for assumptions indicated that all conditions were normally distributed and met assumptions of sphericity for accuracy and reaction time.

2.2.2.1. Accuracy

A repeated measures ANOVA showed that the accuracy differential score did not differ significantly across the conditions with varied probability of stimulus pair occurrence [F(2, 28)=1.851, p=0.176,] with a medium to large effect size ($\eta^{2=}$0.112). As displayed in Table 6, the accuracy differential score decreased by 8.98

points in the 80% probability condition compared to the 50:50 condition, with a medium effect size. The accuracy differential score decreased by 0.56 points in the 20% probability condition compared to the 50:50 condition, with a negligible effect size.

Table 6. Pairwise comparisons for accuracy differential scores between conditions for FM group (n = 15)

Condition (% Accuracy; SD)	**Compared With Condition (% Accuracy; SD)**	**Outcomes (p value; effect size; Confidence Interval)**
50:50 (13.66; 23.53)	20% Probability (13.06; 24.36)	p = 1; d=0.02; 95 CI [-0.69 – 0.74]
	80% Probability (4.67; 15.57)	p = 0.218; d=0.45; 95 CI [-0.2744 - 1.175]

2.2.2.2. Reaction Times

A repeated measures ANOVA showed that the reaction time differential score did not differ significantly across the conditions with varied probability of stimulus pair occurrence [F(2, 28)=1.123, p=0.339] with a medium effect size ($\eta^{2=}$0.074). As outlined in Table 7 the reaction time differential score increased by 0.014 points in the 80% probability condition compared to the 50:50 condition, with a small to medium effect size. The reaction time differential score decreased by 0.08 points in the 20% probability condition compared to the 50:50 condition, with a small effect size.

Table 7. Pairwise comparisons for reaction time differential scores between conditions for FM group (n = 15)

Condition (Mean reaction time; SD)	Compared With Condition (Mean reaction time; SD)	Outcomes (p value; effect size; Confidence Interval)
50:50 (-0.041; 0.032)	20% Probability (-0.05; 0.06)	p = 1; d=0.16; 95 CI [-0.55 – 0.89]
	80% Probability (-0.03; 0.04)	p = 0.911; d=0.38; 95 CI [-0.31 – 1.1]

2.2.3. HC Group ANOVA

Initial tests for assumptions indicated that all conditions were normally distributed and met assumptions of sphericity for accuracy and reaction time.

2.2.3.1. Accuracy

A repeated measures ANOVA with sphericity assumed showed that the accuracy differential score did not differ significantly across the conditions with varied probability of stimulus pair occurrence [F(2, 28)=1.451, p=0.252] with a medium to large effect size ($\eta^{2=}$0.094). As displayed in Table 8 the accuracy differential score decreased by 8.859 points in the 80% probability condition, compared to the 50:50 condition, with a medium to large effect size. The accuracy differential score decreased by 2.77 points in the 20% probability condition, compared to the 50:50 condition, with a small effect size.

Table 8. Pairwise comparisons for accuracy differential scores between conditions for FM group (n = 15)

Condition (% Accuracy; SD)	Compared With Condition (% Accuracy; SD)	Outcomes (p value; effect size; Confidence Interval)
50:50 (11.86; 15.54)	20% Probability (9.09; 27.35)	p = 1; d=0.12; 95 CI [-0.59 – 0.84]
	80% Probability (2.99; 10.75)	p = 0.102; d=0.66; 95 CI [-0.072 – 1.398]

2.2.3.2. Reaction times

A repeated measures ANOVA with sphericity assumed showed that the reaction time differential score did not differ significantly across the conditions with varied probability of stimulus pair occurrence [F(2, 28)=1.995, p=0.155], with a medium to large effect size ($\eta^{2=}$0.125). As displayed in Table 9 the reaction time differential score increased by 0.013 points in the 80% probability condition compared to the 50:50 condition, with a medium effect size. The reaction time differential score increased by 0.005 points in the 20% probability condition compared to the 50:50 condition, with a small effect size.

Table 9. Pairwise comparisons for reaction time differential scores between conditions for HC group (n = 15)

Condition (Mean reaction time; SD)	Compared With Condition (Mean reaction time; SD)	Outcomes (p value; effect size; Confidence Interval)
50:50 (-0.03; 0.03)	20% Probability (-0.03; 0.05)	p = 1; d=0.13; 95 CI [-0.59 – 0.84]
	80% Probability (-0.01; 0.03)	p = 0.22; d=0.45; 95 CI [-0.28 – 1.17]

2.3. To test hypothesis 2, regarding between-groups analysis of WM capacity. Descriptive statistics for between-group differences, with further analysis of effect size using Cohen's D.

The results in Table 10 suggest that the FM group performed more poorly on clinical and experimental measures of WM than HCs in this experiment. It should be noted that the FM group were still in the average range overall for the WAIS-IV WMI, which has a normative mean of 100 and SD of 15.

Table 10. Differences in WAIS-IV WM and n-back scores between FM and HC groups

	FM Group (n = 15)	HC Group (n = 15)	Group difference (t-test p value; effect size (d); Confidence Interval)
Mean WAIS-IV WMI Score (SD)	97.06 (14.62)	104.73 (16.89)	p = 0.227; d=0.49; 95% CI [-0.23 -1.2014]
Mean % accuracy n-back Score (SD)	73.27 (14.57)	81.2 (15.17)	p = 0.155; d=0.53; 95% CI [-0.19 -1.26]

2.4. To test hypothesis 3, regarding variance in AWM which is explained by WM performance. Performance across experimental conditions and groups, and within each group in terms of reaction times on experimental tasks measuring AWM was correlated with performance on the WAIS-IV WMI and n-back task, which measured WM. A further analysis of the correlation between the difference between accuracy differential scores in the 50:50 and 80% probability conditions was also calculated as this variation is hypothesised to be related to AWM. Descriptive statistics and effect sizes are presented as preliminary analysis.

As presented in Table 11, a Pearson's product moment correlation between overall AWM task mean reaction times and percentage accuracy on the 2-back task showed a negligible effect size for both groups combined, a small effect in the FM group and a negligible effect in the HC group. A Pearson's product

moment correlation between overall AWM task mean reaction times and WAIS-IV WMI score for both groups combined showed a negligible effect for both groups combined, a negligible effect in the FM group, and a small effect in the HC group. Similar results were found for the correlation analysis of accuracy differential score differences between the 50:50 and 80% conditions (where 80% probability condition differential scores were subtracted from 50:50 condition differential scores for each participant). Negligible or small effect sizes were found across FM, HC and combined group analysis, for associations between AWM accuracy differential score differences and both the 2-back and WAIS-IV WMI. These findings suggest that performance on the AWM task was only weakly associated with WM performance and there is a level of independence between AWM and WM performance.

Table 11. Correlation analysis between AWM reaction times and WM measures

	2-back Task			**WAIS-IV WMI**		
	FM	**HC**	**Combined**	**FM**	**HC**	**Combined**
AWM Task Mean Reaction Time	p = 0.598 r(13) = 0.15	p = 0.909 r(13) = -0.03	p = 0.982 r(28) = 0.04	p = 0.929 r(13) = 0.03	p = 0.35 r(13) = 0.26	p = 0.843 r(28) = 0.04
AWM Difference Score Between Accuracy Differential Scores in 50:50 and 80% Probability Conditions	p = 0.557 r(13) = 0.17	p = 0.904 r(13) = -0.034	p = 0.712 r(28) = 0.04	p = 0.66 r(13) = -0.12	p = 0.909 r(13) = -0.032	p = 0.48 r(28) = -0.086

2.5. To test hypothesis 4, regarding validation of experimental measure of general WM capacity. Consistency between experimental (2-back task) and clinical (WAIS-IV WMI) measures of WM was analysed using regression analysis (Pearson's product moment correlation coefficient).

A Pearson's product moment correlation coefficient between percentage accuracy on the 2-back task and WAIS-IV WMI score showed a large effect $r(28)=0.58$, $p<0.001$. When 0-back task percentage accuracy was

included as a covariate to control for individual differences in sustained attention capacity, this correlation coefficient still showed a large effect $r(28)=0.52$, $p<0.01$.

2.6. Correlation between measures of depression and WM

Although not one of this study's primary aims, a correlation between PHQ-9 and both WAIS-IV WMI index (auditory-verbal WM measure) and 2-back accuracy scores (visual WM measure) were calculated to add further evidence to the previously reported systematic review. This analysis found that in the FM group (n=16), the correlation coefficient showed a medium negative effect between the PHQ-9 and WAIS-IV WMI $r(14)=-0.34$, $p=0.194$ and a medium negative effect between the PHQ-9 and 2-back $r(14)=0.28$, $p=0.3$. This analysis was not sufficiently powered to detect statistical significance for a medium effect size, but the effect sizes and directionality suggest a relationship between depression and WM in FM, where increased depression scores are associated with decreased WM scores. Interestingly, considering the findings from the systematic review, there was little difference in the effect size when correlating auditory-verbal or visual measures of WM with depression.

Discussion

This study had a primary aim (1) of developing and validating a novel measure of 'top-down' AWM by measuring its external population validity and internal construct validity. It had further aims to (2) provide preliminary data regarding between-groups differences in AWM capacity and the relationship between established measures of general WM and the novel measure of longer-term AWM, which would provide evidence towards concurrent validity of the novel measure and (3) provide preliminary evidence regarding the association between clinical and experimental measures of WM. The results

suggested that there was some evidence that top-down cognitive processes did influence parametric memory processes across both groups combined, thus supporting hypothesis one and providing evidence that the novel measure has construct validity. There was further evidence that the findings of 'top-down' cognitive processes influencing parametric memory processes were consistent across groups using group means and effect sizes, further supporting hypothesis one and indicating that the measure has good population validity. The results suggested that there was a negligible association between the novel measure of AWM and the established measures of general WM, against the predictions of hypothesis 3, which suggested some issues with regards to the concurrent validity of the novel measure, which are discussed below. There was further evidence that the WAIS-IV WMI and n-back task were positively correlated, supporting hypothesis 4 , and that there was some preliminary evidence of differences between experimental and control groups in WM and AWM capacity, which supported hypothesis 2.

The findings fit with past research to an extent. With regard to the standard TOE which has been frequently described in the literature (Harrison et al., 2017), the current study consistently produced an auditory version of this phenomenon, across both groups separately and combined. While this was not the novel aspect central to the measure developed for this study, it was important to demonstrate that the baseline aspect of the measure produced outcomes consistent with extant TOE literature. This was to ensure that the WM processes involved in the TOE were at least evoked in the conventional manner before evaluating whether the extensions developed herein could be considered comparable to this baseline. In addition, there was preliminary evidence that participants with FM scored more poorly on clinical and experimental measures of WM than HCs in line with the literature (Glass, 2008), with medium to large effect sizes found. As frequently reported (Berryman et al., 2013), WM is an area of cognitive difficulty which is often found in FM populations, so while the current study was not powered

to detect statistical significance for the between-groups analysis, this preliminary outcome was expected. It is also important to note that this evidence of reduced general WM capacity in the FM group was quite subtle, which also fits with the existing literature. As even subtle deficits in WM can lead to significant functional difficulties, these findings are potentially clinically relevant. While this study was not powered for robust between-groups analysis, ratings of fatigue, depression and anxiety were elevated in the FM group which is consistent with a large body of literature which indicates that these difficulties are often secondary symptoms occurring in the context of CP including FM. These difficulties may have further contributed to the relative deficits on the general measures of WM. There was also preliminary evidence of increased difficulties for participants with FM compared to HCs on the AWM task, with longer reaction times which may indicate it was more effortful. As this finding related to the novel measure, this study extended the current literature by producing evidence that participants with FM have some subtle difficulties with longer-term AWM, which as the introduction outlined, is not a dimension of WM which has been studied previously. Considering the young literature base characterising the precise nature of WM difficulties which have been widely reported in FM, this is a valuable finding which suggests that one dimension of WM which is compromised in FM is longer timescales. As discussed further in this section, this can inform clinical practice if further research confirms these primary findings.

While the outcomes summarised above fit with the literature, the main hypothesised outcome for the study did not completely transpire in terms of statistical significance, which was that blocks with varying compositions of stimulus train frequencies would produce differing response patterns across groups separately and combined. This was because while there was a significant difference between conditions in the combined FM and HC group, there was no significant finding in each of those groups separately. This was a novel task, so this lack of consistent statistically significant outcomes does not directly

contradict the literature. However, it was based on the general TOE literature, in particular the concept that the human brain comes to anticipate the mean of four stimulus frequencies. As such, the failure to produce statistically significant outcomes across the different groups separately with regard to the anticipated perceptual biases when the mean of four stimulus frequencies was manipulated across conditions was unexpected. The lack of statistically significant difference and negligible effect size when comparing the 50:50 and 20% probability conditions was particularly unexpected.

Despite failing to reach statistical significance across all of the FM, HC and combined groups, there was a significant difference between the 50:50 and 80% probability conditions for the combined group for task accuracy, in the expected direction. This suggests that the novel measure developed for this study was partially able to produce different outcomes across conditions where the probability of stimulus train frequencies was manipulated. As such, it can be concluded that when participants across both groups combined were exposed to stimulus train pairs with 80% frequency, there was a statistically significant reduction in the TOE compared to the standard 50:50 TOE. This suggests an influence of top-down AWM processes across 10s of seconds to minutes on parametric AWM processes as hypothesised.

However, the lack of a significant difference between the 20% probability condition and the 50:50 condition does not directly support the idea that top-down AWM processes influenced automatic AWM processes, particularly as the means of the two conditions suggest that there was a slight reduction in TOE magnitude compared to the standard TOE, against predictions. However it does not necessarily contradict the theory, as there are the following factors to consider. It is possible that the lack of difference between 50:50 and 20% probability conditions is due to the low number of trials in the 20% probability condition, with only four per block, and only two trials per block in which it was hypothesised that discrimination ability would be negatively affected by top-down AWM processes. This means that when participants found it hard to discriminate between some stimulus train pairs, they may be able to score relatively high by chance, meaning that there was a potentially large impact of random noise. The

low number of trials in the 20% probability condition was chosen to reduce the load on participants, both for their well-being and also to minimise the impact of attentional or cognitive fatigue effects on performance. In future studies, a higher number of trials in the 20% probability condition may be required to produce the TOE variation as hypothesised, if one exists.

An alternative explanation for the failure to find an increased TOE in the 20% probability condition is that the mean frequency across all four frequencies in these blocks moved too far from the frequency pairs which occurred with 20% frequency. This could have meant that, in those frequency pairs, as the frequency which was closest to the block mean was further away from the mean than it was in the 50:50 condition, its relative change in comparison to the frequency which was furthest away from the mean was reduced. As such, the perceived difference between the two frequencies in these pairs may not have been as pronounced as hypothesised. For example, in the 80:20 (fast pair: slow pair) ratio blocks, the mean of the frequencies over the block was 24Hz, in comparison to 22.5Hz in the 50:50 ratio blocks. The composition of the trial frequencies remained the same across these blocks, for instance the slow pair may have consisted of 19Hz and 21Hz in both the 50:50 and 80:20 ratio blocks. While the 19Hz stimulus was now further from the overall mean of 24Hz, the 21Hz train was also further from 24Hz, so may not have produced enough of an exaggerated perceived difference compared to the 19Hz train for the hypothesised effect.

It should be noted that the effect size found for all of the group analyses was moderate, and the power calculation for this study was based on a large effect size. As such, it is possible that the significant outcome in the combined group was due to the larger sample being able to detect a smaller effect size than anticipated. As there was no significant interaction of group and experimental condition, the groups, while clinically distinct, did not differ substantially in terms of their outcomes on the longer-term AWM novel measure. While a preliminary outcome, this may suggest that the findings of the combined group are indicative of what the outcomes might be if each group was doubled in size. Indeed, the initial

power calculation indicated that a group of 30 would be sufficient to detect a medium effect size. As such, larger sample sizes for the FM and HC groups may reach statistically significant differences between the conditions. The literature used to determine the sample size for this study concerned general WM, but the effect sizes found in this study can be used to determine more accurate sample sizes for AWM specifically in future studies.

Considering the potential impact of sample size on the study findings, it is worth considering whether there were any consistent findings with regards to effect sizes and condition means across the groups separately and combined. All of the ANOVAs conducted indicated a medium effect size of the condition for the groups separately and combined. Post-hoc analysis indicated that there was generally a medium effect size when comparing the 50:50 condition to the 80% probability condition on measures of both accuracy and reaction time differential scores. The condition means across groups and combined were consistent in showing that there were decreased TOE effects in terms of both reaction time and accuracy in the 80% condition compared to the 50:50 condition. There was also consistent evidence that there were negligible or only small effect sizes when considering differences in accuracy and reaction time differential scores between the 20% probability and 50:50 conditions. The means of the 50:50 and 20% probability conditions suggested that any small differences were either away from the hypothesised direction or only slightly in the direction predicted. These consistent trends suggest that the novel measure of AWM is producing and detecting replicable results across groups which are partly in line with expectations. As such, the findings have good external population validity with regards to the hypothesised outcomes which transpired, while not detecting the full extent of the expected outcomes. As the findings were partly in line with the hypothesised outcomes for the novel measure, it can be said that there is some evidence towards its construct validity, although alterations such as those described above for the 20% condition may be beneficial to confirm this in future studies.

While evidence of population and construct validity is encouraging, it is important to note that the attempt to detect preliminary evidence of concurrent validity was not successful. While this may indeed indicate that the measure does not have good concurrent validity, it should be noted that the novel measure was developed specifically because there is no current measure of WM on a scale of 10s of seconds to minutes. As such, the tests which were used to correlate the novel measure against by definition assess different constructs, as they measure WM on a scale of a few seconds. It was expected that there would be some correlation as these constructs are inter-dependent, but the lack of concurrent validity found here may reflect the fact that these constructs are more distinct than hypothesised. This could be argued to further indicate the need for the novel measure developed herein, as it provides a means of measuring a distinct construct which is not currently covered by existing tests.

Although there was no evidence of a mean difference in top-down effects on AWM between groups, there was a high degree of variability within the FM group, which may indicate individual differences that would be of interest for further study to identify if they have clinical implications. Standard interventions for FM don't routinely offer neuropsychological assessment or intervention for cognitive deficits (Wilson, 2017). Other contexts for clients with cognitive difficulties offer cognitive remediation with good outcomes for cognitive functioning including WM in patients with schizophrenia (Wykes et al., 2018), with additional benefits for wider outcomes such as: improved self-esteem in schizophrenia patients (Wykes et al., 2018); reduced anger in people with a diagnosis of ADHD (Stevenson et al., 2002); reduced anxiety and depression and improved quality of life in people who have suffered an acquired brain injury (Tiersky et al., 2005; Bjorkdahl et al., 2013). In other contexts where cognitive deficits are commonplace, bespoke intervention is guided by appropriate neuropsychological assessments, which have been validated for use in these populations and is guided by the literature on cognitive deficits in these populations (Podell et al., 2010). Developing cognitive measures which are validated in the FM

population and based on robust literature would therefore be expected to improve assessment and intervention in pain management services.

The current study has developed the first measure of AWM on a scale of 10s of seconds to minutes, and demonstrated that it has good external validity with regards to population generalisability, as well as evidence towards establishing construct validity. The current study is early-stage research, as it attempts to develop a measure for a cognitive process which has until now not been researched or assessed clinically in FM. As such, although the establishment of good population validity and indications of construct validity are important steps, the measure requires further steps in terms of validation and also establishing reliability before it can be widely implemented in research or clinical settings. If further research does build on the validity demonstrated in this study, and further validity and reliability is also established for this measure, it would open up possibilities for clinical research and practice. As there is currently a lack of specificity concerning the nature of WM deficits often found in FM, the initial purpose of this measure would be to contribute to clinical research, by enabling investigation of which timescale of WM is affected in FM. This novel measure is of WM on a longer timescale than existing measures of WM, involving different brain regions and levels of cognitive complexity. As such, this measure would be able to more precisely characterise the profile of WM deficits in FM, which is a key area of cognitive difficulty for this population, and may ultimately inform clinical practice.

In clinical settings for other populations with cognitive deficits, direct cognitive rehabilitation treatments have been successful for improving WM capacity for people after a stroke or Traumatic Brain Injury (TBI; Elliott & Parente, 2013), where cognitive function is targeted through exercises and training. Compensatory strategies for WM deficits are also widely used to support WM, for instance learning to 'chunk' information into smaller components improves WM in samples with Alzheimers' disease (Huntley et al., 2011) and modifying or simplifying the environment after a TBI can support WM (Headway, 2009). Considering the evidence of cognitive difficulties, particularly with WM, in FM,

interventions such as these may also be beneficial. If further research using the measure developed in this study can establish that the WM difficulties frequently reported by people with FM are at a relatively long timescale in comparison to controls, clinicians would be able to develop interventions which support WM at this longer timescale. Such interventions could focus on strategies which have been demonstrated to support longer-timescale, top-down WM in other clinical populations, for example, learning to connect information into more meaningful chunks, which has been demonstrated to support retention of longer strings of information in WM over a longer time by exploiting wider cognitive processes (Baddeley, 2000). Conversely, if research using this novel measure does not find evidence of WM difficulties at a longer timescale for people with FM, clinicians might instead focus on interventions for WM deficits at a shorter timescale. This might involve environmental management to minimise cognitive load, or breaking information down into smaller components to facilitate early stage information encoding (Wilson, 2018). Although there has not been any research into the effects of intervention for WM deficits in CP on wider emotional wellbeing and daily functioning, improvements in WM have been found to improve engagement in activities of daily living and return to work in mild TBI (Vallat-Azouvi et al., 2009). It could therefore be anticipated that intervention which supports WM appropriately in FM would have benefits for their quality of life. As the precise characterisation of WM deficits widely reported in FM is currently unclear, it would be valuable for clinical practice to understand this profile in more detail, to make service provision more efficient by guiding intervention more accurately. If further evidence of validity and reliability is established, the measure developed in this study will contribute to research which will refine the evidence base, with implications for clinical practice and client wellbeing as described. The finding of good population validity and construct validity is an important first step towards this end goal.

While the measure will initially benefit clinical research, it could also be used as an assessment in clinical settings ultimately. Neuropsychological assessments are used by Clinical Psychologists across a wide

range of settings, to understand the nature of individual clients' cognitive difficulties and to inform bespoke interventions (Harvey, 2012). As current measures of WM used in clinical settings assess WM on a timescale of a few seconds, but it is known that WM operates up to a timescale of minutes, there is currently a gap in clinical assessment tools for this type of WM. If the measure developed in this study is subsequently demonstrated to have further evidence of internal validity, external validity and reliability, normative data could be developed to provide reference points for mean TOE variation and standard deviations. As with other neuropsychological assessments, this would enable assessment of individuals in clinical settings for WM deficits on a timescale of 10s of seconds to minutes, permitting more personalised interventions (Harvey, 2012). The measure developed herein takes between 30-50 minutes to complete, in line with many existing neuropsychological assessments, and would only require a computer with an internet connection. A number of neuropsychological assessments have been adapted for computerised use recently, so this novel measure would fit with existing methods already employed by clinicians.

Furthermore, uncertainty about the causes of their condition is often reported by clients with FM (Reich et al., 2006), which can cause psychological distress (Mast, 1995). Understanding the neuropsychological profile of FM in more detail would provide clients with greater certainty around their condition with anticipated benefits for their psychological well-being. This study has provided some new insights about the neuropsychological profile of FM. In particular, there is preliminary evidence of a moderate effect size for differences in AWM reaction time. This suggests the new finding that AWM may be more effortful for people with FM than HCs. However, this finding needs to be examined further in larger samples as the above analysis is not sufficiently powered in this study.

Limitations

The study was initially planned to be conducted in a lab setting with researchers present. Due to the COVID-19 pandemic and associated restrictions in the UK, the study was changed to an online project. As such, there were many unknown variables while participants completed the experiment in their own homes. Multiple phone and video calls with each participant were initiated to ensure they had an appropriate set-up and to emphasise the appropriate environment requirements. There were also questionnaires at the end of each task to check participant progress and ask about any distractions. However, the data is likely to be poorer quality than it would have been in a lab setting.

The change to an online study also required participants to be relatively confident in their own computer literacy skills. Several participants withdrew due to anxiety about using computer equipment. Furthermore, it is possible that potential participants with anxiety about their cognition may have avoided the experiment, which may mean that relative difficulties with WM and AWM for people with FM indicated in this report may be an underestimate. However, it may also have been possible that participants with decreased mobility or anxiety about leaving home were more likely to take part in this online version. A more representative sample, including both in-person recruitment from clinical or support group settings when possible (COVID-permitting) and online recruitment, may be more accurate than the sample herein who were only recruited remotely and thus needed greater personal initiative to participate. Also, as the computer experiment took 30-40 minutes, participants with pain or discomfort when sitting may have been put off, despite regular break points integrated into the experiment. As such, the invitation to participate may have been more appealing to a self-selecting group of participants with greater confidence in their computer and cognitive skills, and with less physical discomfort in sitting.

Recommendations for future research

Due to the findings which suggested that the variation in TOE was partially produced when both groups were combined, but not when analysed individually, it is recommended that the measure is tested with a larger sample of both the FM and HC groups. This could make use of the effect sizes found in this study to conduct a new sample size calculation appropriate to AWM. This would elongate the experiment however and would need to be developed with patient groups to ensure that it is acceptable and would not be significantly impacted upon by difficulties with sustained attention or fatigue. Furthermore, the experiment may benefit from being conducted in a lab setting, to minimise distractions or other confounding factors, and standardise the equipment used.

The findings were partly in line with the hypothesised patterns of results, which provided evidence towards internal construct validity, and a repeated measures design was used which reduced the impact of extraneous variables, which would be expected to improve internal validity. However, a next step for this measure would be to further establish internal validity through targeted research. Future studies could aim to provide evidence that the measure outcomes correlate with established measures of WM which have not been attempted in this study for example, which would suggest that the novel measure does have concurrent validity. As the established measures used in this study to investigate concurrent validity of the novel measure assessed WM on a short timescale, it would be important to use one which taps into longer-term WM. Although no other measures exist which assess WM at the same timescale as the novel measure (hence the need for it), there are some which assess WM at a longer timescale than the WAIS-WMI and n-back. For example, complex span tasks have been argued to access WM processes on a longer timescale than other measures including the n-back (Wilhelm et al., 2013), and may be a more appropriate reference for the novel measure. Additionally, future studies could assess whether the measure outcomes accurately predict whether difficulties identified are associated with poorer clinical outcomes which are known to be related to WM, which would evidence predictive validity. As another means of assessing internal validity, asking participants in future studies to rate their

perceptions of whether the novel measure appears to measure longer-term AWM would provide some indication of whether the measure has face validity (Hardesty & Bearden, 2004). It would also be beneficial to establish whether the measure has wider external validity, for example by repeating the measure using the same procedure but in different contexts and at different times.

Conclusion

This study has developed and provided initial evidence of population and construct validity for a novel measure of AWM on a scale of 10s of seconds to minutes for use in FM. In addition, it has provided preliminary evidence that tasks involving AWM may be more effortful for people with FM than HCs, as well as preliminary evidence of a positive correlation between clinical and experimental measures of WM. The novel measure of AWM can be used for assessment in clinical settings and for further research into this topic. The new preliminary findings regarding more effortful AWM in FM need further research with larger samples, before potentially being used to inform interventions and reduce uncertainty in FM populations.

References

Abdel Shaheed, C., Maher, C. G., Williams, K. A., Day, R., & McLachlan, A. J. (2016). Efficacy, tolerability, and dose-dependent effects of opioid analgesics for low back pain: A systematic review and meta-analysis. *JAMA Internal Medicine*, *176*(7), 958–968. https://doi.org/10.1001/jamainternmed.2016.1251

Ambrose, K. R., Gracely, R. H., & Glass, J. M. (2012). Fibromyalgia dyscognition: Concepts and issues. *Reumatismo*, *64*(4), 206–215. https://doi.org/10.4081/reumatismo.2012.206

American Psychological Association (2020). *Patient health questionnaire (PHQ-9 & PHQ-2).* Retrieved from https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/patient-health

Andrade C. (2018). Internal, External, and Ecological Validity in Research Design, Conduct, and Evaluation. *Indian Journal of Psychological Medicine*, *40*(5), 498–499. https://doi.org/10.4103/IJPSYM.IJPSYM_334_18

Aronson, E., Ellsworth, P., Carlsmith, J.M. & Gonzalez, M.H. (1990). M*ethods of Research in Social Psychology.* McGraw-Hill Education, New York.

Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PloS One*, *6*(5), e19551. https://doi.org/10.1371/journal.pone.0019551

Attal, N., Masselin-Dubois, A., Martinez, V., Jayr, C., Albi, A., Fermanian, J., Bouhassira, D., & Baudic, S. (2014). Does cognitive functioning predict chronic pain? Results from a prospective surgical cohort. *Brain : A Journal of Neurology*, *137*(Pt 3), 904–917. https://doi.org/10.1093/brain/awt354

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baliki, M. N., Geha, P. Y., Apkarian, A. V., & Chialvo, D. R. (2008). Beyond feeling: chronic pain hurts the brain, disrupting the default-mode network dynamics. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(6), 1398–1403. https://doi.org/10.1523/JNEUROSCI.4123-07.2008

Berryman, C., Stanton, T. R., Jane Bowering, K., Tabor, A., McFarlane, A., & Lorimer Moseley, G. (2013). Evidence for working memory deficits in CP: a systematic review and meta-analysis. *Pain, 154*(8), 1181–1196. https://doi.org/10.1016/j.pain.2013.03.002

Bertolucci, P. H., & de Oliveira, F. F. (2013). Cognitive impairment in fibromyalgia. *Current Pain and Headache Reports*, *17*(7), 344. https://doi.org/10.1007/s11916-013-0344-9

Bilder R. M. (2011). Neuropsychology 3.0: evidence-based science and practice. *Journal of the International Neuropsychological Society : JINS*, *17*(1), 7–13. https://doi.org/10.1017/S1355617710001396

Binetti, N., Harrison, C., Mareschal, I., & Johnston, A. (2017). Temporal Order Judgements of Dynamic Gaze Stimuli Reveal a Postdictive Prioritisation of Averted Over Direct Shifts. I-*Perception, 8*(4), 2041669517720808. https://doi.org/10.1177/2041669517720808

Björkdahl, A., Akerlund, E., Svensson, S., & Esbjörnsson, E. (2013). A randomized study of computerized working memory training and effects on functioning in everyday life for patients with brain injury. *Brain Injury*, *27*(13-14), 1658–1665. https://doi.org/10.3109/02699052.2013.830196

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Cohen, H., McCabe, C., Harris, N., Hall, J., Lewis, J., & Blake, D. R. (2013). Clinical evidence of parietal cortex dysfunction and correlation with extent of allodynia in CRPS type 1. *European Journal of Pain (London, England)*, *17*(4), 527–538. https://doi.org/10.1002/j.1532-2149.2012.00213.x

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Crombez, G., Eccleston, C., Baeyens, F., & Eelen, P. (1998). When somatic information threatens, catastrophic thinking enhances attentional interference. *Pain*, *75*(2-3), 187–198. https://doi.org/10.1016/s0304-3959(97)00219-4

Davis, E., Young, D., Waters, E., & Gold, L. (2017). Children and adolescents, measuring the quality of life of. In S.R. Qhah (Ed.), *International Encyclopedia of Public Health* (pp. 539-545). Academic Press.

Demirci, S., & Savas, S. (2002). The auditory event related potentials in episodic and chronic pain sufferers. *European Journal of Pain (London, England)*, *6*(3), 239–244. https://doi.org/10.1053/eujp.2001.0342

Elliott, M., & Parente, F. (2014). Efficacy of memory rehabilitation therapy: a meta-analysis of TBI and stroke cognitive rehabilitation literature. *Brain Injury*, *28*(12), 1610–1616. https://doi.org/10.3109/02699052.2014.934921

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037/0033-295x.102.2.211

Fabrigar, L. R., Wegener, D. T., & Petty, R. E. (2020). A validity-based framework for understanding replication in psychology. *Personality and Social Psychology Review, 24*(4), 316–344. https://doi.org/10.1177/1088868320931366

Fassihi, A., Akrami, A., Esmaeili, V., & Diamond, M. E. (2014). Tactile perception and working memory in rats and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(6), 2331–2336. https://doi.org/10.1073/pnas.1315171111

Fuster, J. M., & Bressler, S. L. (2012). Cognit activation: a mechanism enabling temporal integration in working memory. *Trends in Cognitive Sciences*, *16*(4), 207–218. https://doi.org/10.1016/j.tics.2012.03.005

Gajewski, P. D., Hanisch, E., Falkenstein, M., Thönes, S., & Wascher, E. (2018). What does the *n*-back task measure as we get older? relations between working-memory measures and other cognitive functions across the lifespan. *Frontiers in Psychology*, *9*, 2208. https://doi.org/10.3389/fpsyg.2018.02208

Garcia-Alvarez, L., Gomar, J. J., Sousa, A., Garcia-Portilla, M. P., & Goldberg, T. E. (2019). Breadth and depth of working memory and executive function compromises in mild cognitive impairment and their relationships to frontal lobe morphometry and functional competence. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, *11*, 170–179. https://doi.org/10.1016/j.dadm.2018.12.010

Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neuroscience*, *8*(10), 1298–1300. https://doi.org/10.1038/nn1543

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Gil-Ugidos, A., Rodríguez-Salgado, D., Pidal-Miranda, M., Samartin-Veiga, N., Fernández-Prieto, M., & Carrillo-de-la-Peña, M. T. (2021). Working memory performance, pain and associated clinical variables in women with fibromyalgia. *Frontiers in Psychology*, *12*, 1-13. https://doi.org/10.3389/fpsyg.2021.747533

Glass J. M. (2008). Fibromyalgia and cognition. *The Journal of Clinical Psychiatry*, *69 Suppl 2*, 20–24.

Green P. (2007). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America*, *18*(1), 43–vi. https://doi.org/10.1016/j.pmr.2006.11.002

Greiffenstein, M.F., Baker, W., & Gola,T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*(3), 218-224.

Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development: Implications for improving face validity of measures of unobservable constructs. *Journal of Business Research, 57*(2), 98–107. https://doi.org/10.1016/S0148-2963(01)00295-8

Harrison, C., Binetti, N., Mareschal, I., & Johnston, A. (2017). Time-Order Errors in Duration Judgment Are Independent of Spatial Positioning. *Frontiers in Psychology*, *8*, 340. https://doi.org/10.3389/fpsyg.2017.00340

Harvey P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues in Clinical Neuroscience*, *14*(1), 91–99. https://doi.org/10.31887/DCNS.2012.14.1/pharvey

Hays, R.D., & Reeve, B.B. (2008). Measurement and modelling of health-related quality of life. In Killewo (Ed*.), International Encyclopedia of Public* Health (pp.241-252). Academic Press. https://doi.org/10.1016/B978-012373960-5.00336-1

Headway (2009). *Coping with memory problems: practical strategies.* Retrieved from https://www.headway.org.uk/media/8821/coping-with-memory-problems-practical-strategies-factsheet.pdf

Hellström A. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology. Human Perception and Performance, 5(*3), 460–477.

Hellström, Å., & Rammsayer, T. H. (2015). Time-order errors and standard-position effects in duration discrimination: An experimental study and an analysis by the sensation-weighting model. *Attention, Perception & Psychophysics*, *77*(7), 2409–2423. https://doi.org/10.3758/s13414-015-0946-x

Holdnack, J.A. (2019). The development, future and expansion of the WAIS-IV as a cornerstone in comprehensive cognitive assessments. In G. Goldstein, D.N. Allen & J. DeLuca (Eds.), *Handbook of Psychological Assessment Fourth Edition* (pp.103-139). Academic Press.

Huntley, J., Bor, D., Hampshire, A., Owen, A., & Howard, R. (2011). Working memory task performance and chunking in early Alzheimer's disease. *The British Journal of Psychiatry : The Journal of Mental Science*, *198*(5), 398–403. https://doi.org/10.1192/bjp.bp.110.083857

Jadhav, S. P., & Feldman, D. E. (2010). Texture coding in the whisker system. *Current Opinion in Neurobiology*, *20*(3), 313–318. https://doi.org/10.1016/j.conb.2010.02.014

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Legrain, V., Iannetti, G. D., Plaghki, L., & Mouraux, A. (2011). The pain matrix reloaded: a salience detection system for the body. *Progress in Neurobiology*, *93*(1), 111–124. https://doi.org/10.1016/j.pneurobio.2010.10.005

Legrain, V., Damme, S. V., Eccleston, C., Davis, K. D., Seminowicz, D. A., & Crombez, G. (2009). A neurocognitive model of attention to pain: behavioral and neuroimaging evidence. *Pain*, *144*(3), 230–232. https://doi.org/10.1016/j.pain.2009.03.020

Mast, M.E. (1995). Adult uncertainty in illness: A critical review of research. *Scholarly Inquiry for Nursing Practice: An International Journal.*, *9*(1), 3–24.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222–243. https://doi.org/10.1037/a0017619

McDermott, R. (2011*).* Internal and external validity. In J.N. Druckman, D.P. Green, J.H. Kuklinski & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 27-40). Cambridge.

Moore, D. J., Meints, S. M., Lazaridou, A., Johnson, D., Franceschelli, O., Cornelius, M., Schreiber, K., & Edwards, R. R. (2019). The effect of induced and chronic pain on attention. *The Journal of Pain*, *20*(11), 1353–1361. https://doi.org/10.1016/j.jpain.2019.05.004

National Health Service (2019). *Fibromyalgia: Overview*. Retrieved from https://www.nhs.uk/conditions/FM/

Pain Proposing Steering Committee (2010). *Pain proposal: Improving the current and future management of chronic pain. A European consensus report.* Retrieved from https://europeanpainfederation.eu/wp-content/uploads/2016/06/pain_proposal.pdf

Patino, C. M., & Ferreira, J. C. (2018). Internal and external validity: can you apply research study results to your patients?. *Jornal Brasileiro de Pneumologia : Publicacao Oficial da Sociedade Brasileira de Pneumologia e Tisilogia*, *44*(3), 183. https://doi.org/10.1590/S1806-37562018000000164

Perini, G., Cotta Ramusino, M., Sinforiani, E., Bernini, S., Petrachi, R., & Costa, A. (2019). Cognitive impairment in depression: recent advances and novel treatments. *Neuropsychiatric Disease and Treatment*, *15*, 1249–1258. https://doi.org/10.2147/NDT.S199746

Phillips C. J. (2009). The Cost and Burden of CP. *Reviews in Pain*, *3*(1), 2–5. https://doi.org/10.1177/204946370900300102

Podell, K., Gifford, K., Bougakov, D., & Goldberg, E. (2010). Neuropsychological assessment in traumatic brain injury. *The Psychiatric Clinics of North America*, *33*(4), 855–876. https://doi.org/10.1016/j.psc.2010.08.003

Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How recent history affects perception: the normative approach and its heuristic approximation. *PLoS Computational Biology*, *8*(10), e1002731. https://doi.org/10.1371/journal.pcbi.1002731

Reich, J. W., Olmsted, M. E., & van Puymbroeck, C. M. (2006). Illness uncertainty, partner caregiver burden and support, and relationship satisfaction in FM and osteoarthritis patients. *Arthritis and Rheumatism*, *55*(1), 86–93. https://doi.org/10.1002/art.21700

Ricker, T. J., AuBuchon, A. M., & Cowan, N. (2010). Working memory. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(4), 573–585. https://doi.org/10.1002/wcs.50

Romo, R., & Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. *Nature Reviews. Neuroscience*, *4*(3), 203–218. https://doi.org/10.1038/nrn1058

Seo, J., Kim, S. H., Kim, Y. T., Song, H. J., Lee, J. J., Kim, S. H., Han, S. W., Nam, E. J., Kim, S. K., Lee, H. J., Lee, S. J., & Chang, Y. (2012). WM impairment in FM patients associated with altered frontoparietal memory network. *PloS One*, *7*(6), e37808. https://doi.org/10.1371/journal.pone.0037808

Shamosh, N.A., DeYoung, C.G., Green, A.E., Reis, D.L., Johnson, M.R., & Conway, A.R. (2008). Individual differences in delay discounting: relation to intelligence, WM, and anterior prefrontal cortex. *Psychol. Sci., 19,* 904–911.

Silva, M.A. (2008). Development of the WAIS-III: A Brief Overview, History, and Description. *Graduate Journal of Counseling Psychology, 1,* 11.

Spitzer, B., & Blankenburg, F. (2012). Supramodal parametric WM processing in humans. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(10), 3287–3295. https://doi.org/10.1523/JNEUROSCI.5280-11.2012

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stevenson, C. S., Whitmont, S., Bornholt, L., Livesey, D., & Stevenson, R. J. (2002). A cognitive remediation programme for adults with Attention Deficit Hyperactivity Disorder. *The Australian and New Zealand Journal of Psychiatry*, *36*(5), 610–616. https://doi.org/10.1046/j.1440-1614.2002.01052.x

Stites, S. D., Harkins, K., Rubright, J. D., & Karlawish, J. (2018). Relationships Between Cognitive Complaints and Quality of Life in Older Adults With Mild Cognitive Impairment, Mild Alzheimer Disease Dementia, and Normal Cognition. *Alzheimer Disease and Associated Disorders*, *32*(4), 276–283. https://doi.org/10.1097/WAD.0000000000000262

Strauss, M. E., & Smith, G. T. (2009). Construct validity: advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*, 1–25. https://doi.org/10.1146/annurev.clinpsy.032408.153639

Sun, Y., Fu, Z., Bo, Q., Mao, Z., Ma, X., & Wang, C. (2020). The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital. *BMC Psychiatry*, *20*(1), 474. https://doi.org/10.1186/s12888-020-02885-6

Tetzlaff, C., Kolodziejski, C., Markelic, I., & Wörgötter, F. (2012). Time scales of memory, learning, and plasticity. *Biological Cybernetics*, *106*(11-12), 715–726. https://doi.org/10.1007/s00422-012-0529-z

Tiersky, L. A., Anselmi, V., Johnston, M. V., Kurtyka, J., Roosen, E., Schwartz, T., & Deluca, J. (2005). A trial of neuropsychologic rehabilitation in mild-spectrum traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *86*(8), 1565–1574. https://doi.org/10.1016/j.apmr.2005.03.013

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it?. *Frontiers in Psychology*, *4*, 433. https://doi.org/10.3389/fpsyg.2013.00433

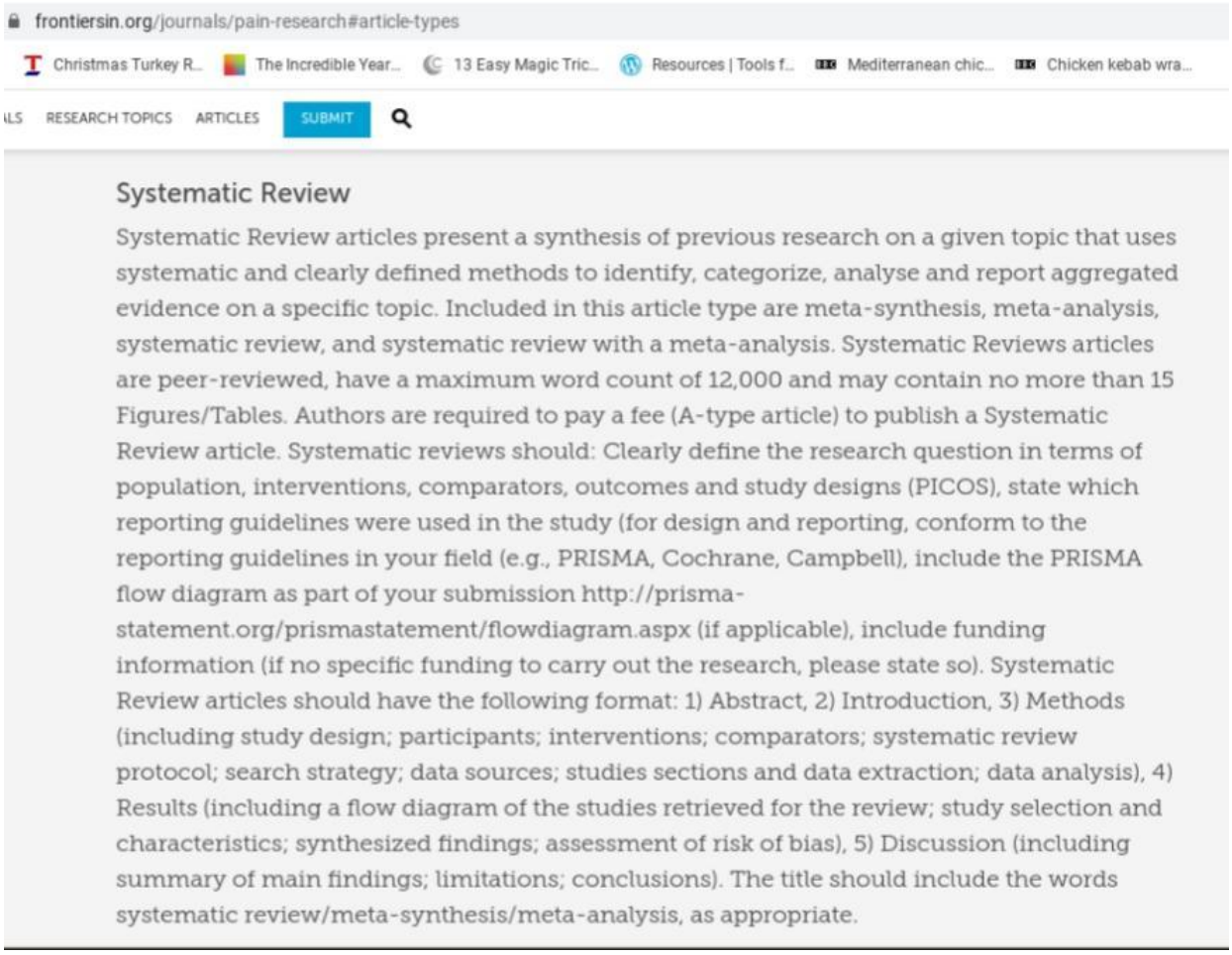
Wilson, B.A. (2018). *Memory problems after brain injury.* Retrieved from https://www.headway.org.uk/media/8777/memory-problems-after-brain-injury-e-booklet.pdf

Wilson I. R. (2017). Management of CP through pain management programmes. *British Medical Bulletin*, *124*(1), 55–64. https://doi.org/10.1093/bmb/ldx032

Wykes, T., Reeder, C., Landau, S., Everitt, B., Knapp, M., Patel, A., & Romeo, R. (2007). Cognitive remediation therapy in schizophrenia: Randomised controlled trial. *British Journal of Psychiatry, 190*(5), 421-427. https://10.1192/bjp.bp.106.026575

Appendix A

Author Guidelines for Frontiers in Pain Research Systematic Reviews



Detailed author guidelines are available here: https://www.frontiersin.org/about/author-guidelines

Appendix B

Standardised Data Extraction Form

Data Extraction Table

Descriptive Data List	Descriptive Data Extracted	Analytical Data List	Analytic Data Extracted
First Author's name	Baker, K.S	Measures used (depression/ working memory)	BDI (depression); BRIEF-A self-report measure (WM)
		Tests administered by	
Year of publication	2017	Working memory outcome e.g. standard deviation/ percentile on a neuropsychological assessment or reaction time/ accuracy on a computerised task.	BRIEF-A: T scores, based on age-adjusted normative data. Clinical cut-off over 65
Full text paper or abstract?	Full Paper	Depression outcome	Total scale score (and clinical categories)
		Type of analysis	Partial correlation (spearman's r)
Setting (clinical/ lab/ other?)	Not specified	Number of participants in analysis	39
Study design	Cross-sectional	Significance (p value)	False discovery rate adjusted p value <0.05 (significant)
		Effect size	0.50
Number of participants with CP (including dropouts)	41 started, 2 dropped out		Conclusion: significant positive correlation between depression and WM
Study sponsorship	The project was supported by doctoral project funding to KSB from the School of Psychological Sciences and Monash Institute for Cognitive and Clinical Neurosciences, Monash University.		
CP participant	Mean age: 42.97, range		

Appendix C

Modified Newcastle-Ottowa Quality Assessment Scale

Modified Newcastle-Ottawa Quality Assessment Scale (NOS)

A study can be awarded a maximum of one star for each numbered item within the Selection, Exposure, and Outcome categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the sample/case definition adequate? – chronic pain status

a) Yes, with independent validation * - assessed before/during the study according to some established criteria

b) Yes, e.g. record linkage (medical charts) or based on self-reports

c) No description

2) Representativeness of the sample/exposed cohort

a) Truly representative of the average in the target population *- random sampling or consecutive recruitment, or discussion of sample characteristics in relation to whether they're similar to those reported in other studies looking at the same condition

b) Somewhat representative of the average in the target population * - as above but not completely; if cognitive complaints constitute inclusion criteria this would be okay as widely reported in the population

c) Selected group of users e.g. nurses, volunteers –other sample characteristics that make it very specific/not representative

d) No description of the sampling strategy

3) Selection of the controls/non-exposed cohort – if there's no healthy control group, consider normative data for cognitive assessments as an equivalent

a) Drawn from the same community as the sample/exposed cohort * some indication of both cultural and demographic similarity between normative/control sample and patients

b) Drawn from a different source – if the above criteria not met

c) No description of the derivation of the non-exposed cohort

4) Definition of controls

a) No history of disease/exposure * - if considering normative data, assume that it comes prom participants without chronic pain

b) No description of source

Comparability

5) Comparability of sample and control groups on the basis of the design or analysis – covariates in an ANOVA or multivariate analyses like multiple regression, or other way of controlling for confounders (original NOS guidelines state that lack of significant differences between groups on baseline characteristics is not sufficient for a star on this item);

a) Study controls for AGE (select the most important factor) *

b) Study controls for any additional factor * education level, depression, anxiety, medication, fatigue

Exposure

6) Ascertainment of exposure –independent variable(s) as stated in the study aims/hypotheses

a) Validated measurement tool * - if depression is considered, this would likely be validated as already stated in the review inclusion criteria

b) Structured clinical interview *

c) Written self-report/non-validated – if validated self report measures were used, choose (a)

d) No description

7) Same method of ascertainment for sample/case and controls

a) Yes *

b) No

8) Non-response rate and description –any exclusions of participants once they've been enrolled in the study

a) Same rate for both groups and clearly described *

b) Non-respondents described

c) Rate different and no designation – if there is no mention of any exclusions (but if numbers analyzed match the numbers recruited, this would indicate complete data)

Outcome

9) Assessment of outcome –likely something related to cognitive functioning

a) Validated measurement tool * - most likely because validated measures should be used for a study to be included in the review

b) Structured clinical interview

c) Written self-report/non-validated –justified to choose (a) if a measure is self-reported but validated

d) No description

Total = / 10

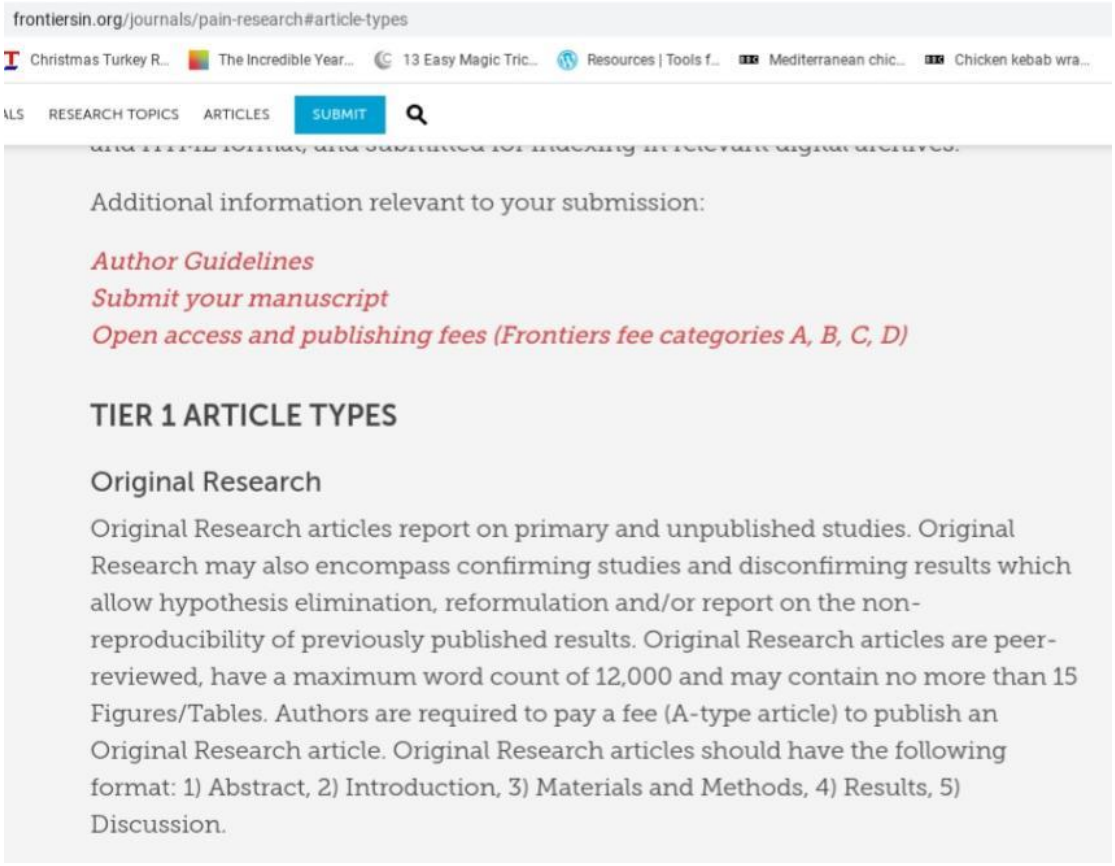
This scale has been adapted from the Newcastle-Ottawa Quality Assessment Scale (NOS; Wells et al., 2014) for cohort and case-control studies to perform a quality assessment for this systematic review.

Appendix D
Quality Assessment of Included Studies

Criteria	*Acceptable (*)*	*Baker et al. (2017)*	*Baldanok et al. (1999)*	*Dick et al. (2002)*	*Elvenia et al. (2015)*	*Grossi et al. (2018)*	*Herbert et al. (2018)*	*Latyshova et al. (2019)*	*Oncu et al. (2015)*	*Pidal-Miranda et al. (2018)*	*Torkamani et al. (2015)*
					Selection						
Is the Case Definition Adequate?	Yes, with independent validation.	-	-	*	-	*	*	*	*	*	*
Representativeness of Cases	Truly/ somewhat representative of the average in the target population	*	-	*	-	*	-	*	-	-	-
Selection of Controls	Drawn from the same community	-	-	*	*	-	*	*	-	*	*
Definition of controls	No history of disease	*	-	*	*	-	*	*	-	*	*
					Comparability						
Study controls for Age	Yes	*	-	*	-	-	-	*	*	-	-
Study controls for any of education level, anxiety, medication, fatigue and depression		*	-	*	*	*	*	*	*	*	-
					Exposure						
Ascertainment of exposure	Validated measurement tool.	*	*	*	*	*	*	*	*	*	*
Same Method of Ascertainment for Cases and Controls	Yes	-	-	-	*	*	*	*	*	*	*
Non-Response Rate	Same rate for both groups	*	-	*	-	*	*	*	-	*	*
Assessment of Outcome	Validated measurement tool	*	*	*	*	*	*	*	*	*	*
Total Quality Score (Maximum = 10)		**7**	**2**	**9**	**6**	**7**	**8**	**10**	**6**	**8**	**7**

Appendix E

Author Guidelines for Frontiers in Pain Research Original Research



Detailed author guidelines are available here: https://www.frontiersin.org/about/author-guidelines

Appendix F

E-mail invitation for Study Participation

Dear XXXX,

I am writing to invite you to take part in a research study conducted by the University of Liverpool, which you may be eligible for. This study will investigate how brain function contributes to fibromyalgia. The study will be conducted entirely online and you will be reimbursed for your time with an Amazon voucher.

You are receiving this invitation because you have previously taken part in research with Dr Chris Brown and Eleanor Brian from the University of Liverpool. This previous project was a brain imaging study called MoNoPly, following which you expressed an interest in taking part in further research. If you want to participate in the current study, or if you would rather not receive further invitations to participate in research, please let us know by filling in the Expression of Interest form below. This can be returned to me by clicking reply to this message. Alternatively, you can reply to this e-mail without using the Expression of interest form, or contact me by phone on XXXX if you prefer.

Expression of interest form

Please delete as appropriate:

□ I would like to participate in this study, or request further information before deciding

□ I would not like to participate in this study, but would be interested in receiving information about other research studies in the future

□ I would not like to receive information about any research studies and do not consent to my medical records being accessed to contact me about research.

We hope to hear from you.

Regards,

XXXX

Doctoral Trainee Clinical Psychologist, University of Liverpool

e-mail: x.xxxx@liverpool.ac.uk

telephone: XXXX

Appendix G

Participant Information Sheet

PARTICIPANT INFORMATION SHEET FOR PARTICIPANTS WITH CHRONIC PAIN

Title of Project	Auditory Working Memory in Chronic Pain (AWMiCP)
Chief Investigator	Dr Christopher Brown

You are being invited to take part in a research study. Before deciding whether to take part, you need to understand why this research is being done and what it involves. Please take time to read the following information carefully and talk to others about the study if you wish. *Please ask us if anything is not clear or if you would like more information.* Please take your time to decide whether or not you wish to take part.

Section 1 tells you the purpose of the study and what will happen to you if you take part.

Section 2 gives you more detailed information about the conduct of the study.

Section 1: Purpose of the study and what will happen

What is the purpose of the study?

Chronic pain (i.e. pain lasting more than 3 months) can persist despite the best efforts of doctors. For many patients, treatments such as pain-killing medications simply don't work well, and not everyone can be successfully treated with surgery. This may be because these treatments are not targeting all of the causes of the pain. The feeling of pain depends on complex processes in the nervous system including in the brain. Research has found that how the brain makes sense of sound sensations in the moment they happen and straight afterwards (termed Auditory Working Memory) is related to chronic pain symptoms, and might even explain why some people develop chronic pain symptoms in the first place. But we don't understand very well what causes this to happen or how to treat it.

With this project, we plan to measure the ability of the brain to make sense of sound sensations, and then to use this information to analyse what aspects of brain function might be contributing to pain. We anticipate that in the future, the information from this project will be used to find more effective ways of treating chronic pain.

Why have I been invited?

You have been invited because you have previously sought medical treatment for fibromyalgia. We are including participants in this study who are over the age of 18 and under the age of 65.

Do I have to take part?

No – participating in this study is completely voluntary. If you decide to participate you will be asked to sign an Informed Consent Form, however you are still free to change your mind and leave the study at any time (including before or during the study) without needing to give a reason. See Section 2 of this form for further details.

Are there any reasons why I cannot take part?

We will discuss with you in detail whether there are any reasons why it would be unsuitable for you to take part in this study. In summary, you will **not** be able to take part if you:

- Are currently or plan to be hospitalised during the period of study.
- Have a history of serious head injury or brain surgery.
- Have a history of neurological disease.
- Have a diagnosis of a developmental condition such as Intellectual Disability or Autistic Spectrum Disorder.
- Are not able to follow the study procedure for other reasons not identified above.

What will happen to me if I take part?

Before we ask you to take part in the study, we will go over the study with you on the phone to ensure that you are eligible and that you are happy to continue.

The study will take place in your own home and would last approximately 1 hour (including some breaks for rest or refreshments). You do not have to stay for the full duration if you do not wish to, and can leave any time you like without providing a reason. At the start, you will be asked to sign an Informed Consent Form to say that you are happy to undergo the study procedures - you will be given a copy of this to take away and refer to later.

Here is an overview of the study visit and what will be involved specifically. The section following this overview contains more detail on the study procedures (i.e. those we indicate here in underlined *italics*) should you wish to know more.

Visit

You will be invited to join a video call appointment with a researcher from the University of Liverpool. Initially, you will be asked to complete some *question-and-answer tests* verbally by video call which assess your general memory. You will then be asked to complete a *computer test* measuring your short-term memory. Following this, you will be asked to undergo some tasks that involve hearing *sound sensations* that will be played to you through your computer speaker or headphones. These sensations will not be uncomfortable and you will be able to adjust the volume of the sound to ensure you are comfortable. We will then measure how your brain responds to these sensations by asking you to make judgements about their speed and frequency. We will also ask you to complete a number of online questionnaires to be completed at home. These questionnaires measure your chronic pain symptoms,

your ability to perform normal daily activities, and how well you cope with the pain mentally and physically.

Further detail on study procedures

Verbal Tests

This study is examining the relationship between chronic pain and holding auditory information in mind for short periods (termed Auditory Working Memory). The ability to hold auditory information in mind for short periods is influenced by the more general ability to hold any type of information in mind for short period (termed Working Memory). We will conduct some widely used verbal tests of Working Memory to make sure that when we measure your Auditory Working Memory, we can adjust the scores for individual differences. This will involve repeating information which the researcher says and doing some calculations in your mind. These tests will not have your name on them, and will be anonymous upon completion.

Computer Test

Similarly to the paper-based tests, we aim to make this study thorough by adjusting for general differences in Working Memory. A widely used computer task will be used to gather additional information about your Working Memory. This task is called the N-Back task, and involves remembering and selecting images which were presented a few seconds previously. As with all data gathered in this study, your scores on this test will be anonymous.

Sound sensations

We will conduct tasks in which you will hear sound stimuli through your own computer speakers or headphones (each one lasting no more than a second or two). This is an entirely safe procedure. Initially, we will spend time adjusting the volume of the sound to best suit you. During the tasks, these sensations will be presented in blocks at varying speeds and you will be asked to respond by saying which blocks of sensations were faster.

Are there any restrictions that I need to comply with if I take part?

You will be asked to refrain from consuming alcohol, or consuming any other recreational drugs for 24 hrs prior to the study visits, in case it interferes with task performance.

What are the possible benefits of taking part?

The study does not involve any form of treatment and so you are not expected to benefit directly from participating in this study. However, information collected as part of your participation may benefit patients with chronic pain in the future. All participants will be offered a £10 Amazon voucher as remuneration for their time.

What are the possible disadvantages and risks of taking part?

The study is unlikely to cause risk to your health should you take part. Some things to note are:

- Occasionally, tests of cognition including Working Memory or Auditory Working Memory can be anxiety provoking or frustrating if you find it harder than expected. A mental health professional will be present as part of the research team if you require support, and you are free to leave the study at any point.
- COVID-19: As this study is online and does not require you to leave your home, there is no anticipated risk associated with Covid-19.

What are the costs of taking part?

As this study is online, there are no anticipated significant costs such as travel. However, you will be making a valuable investment of your time, so a £10 Amazon voucher will be offered to all participants in recognition of this.

If the information in Part 1 has interested you and you are considering taking part in the study, please read the additional information in Part 2 before making any decision.

Section 2: Study Conduct

What if I decide I no longer wish to participate in the study?

You are free to withdraw from the study at any time without giving a reason. No questions will be asked and there will not be any ramifications. For example, your future medical treatment will not be affected in any way.

The study doctor may also choose to withdraw you from the study if they feel it is in your best interests or if you have been unable to comply with the requirements of the study.

What if there is a problem?

Minor complaints

If you have a minor complaint then you need to contact the researcher(s) in the first instance. Please contact Dr Christopher Brown at 0151 794 2174 or christopher.brown@liverpool.ac.uk.

Formal Complaints

If you wish to make a formal complaint or if you are not satisfied with the response you have gained from the researchers in the first instance, then please contact the Research Governance Officer on 0151 794 8290 (ethics@liv.ac.uk). When contacting the Research Governance Officer, please provide details of the name or description of the study (so that it can be identified), the researcher involved, and the details of the complaint you wish to make.

Harm

In the event that something does go wrong and you are harmed during the research you may have grounds for a legal action for compensation against the University of Liverpool but you may have to pay your legal costs.

Will my participation in this study be kept confidential?

All information collected about you as a result of your participation in the study will be anonymised and kept strictly confidential. Your personal and medical information will be kept in a secure file and be treated in the strictest confidence. We will not store your contact details after the study ends unless you have given us permission to do so. For example, you may wish us to keep you informed about the results of the research, or opportunities to participate in other studies that are relevant to you. You may ask to see your personal information at any time and correct any errors if necessary. If you wish to view your personal information, please write to the Liverpool University Data Protection Officer, Computing Services Department, Chadwick Building, Peach Street, Liverpool, L69 7ZF, for more information on how to do this.

Once you have agreed to participate in this study you will be allocated a unique study number which will be used on all your study documentation. Once you have completed participation in the study, your study data will only be identifiable by this unique number. Your data will therefore be fully anonymised. "Anonymised" means that your data will be labelled with a code and all personal identifying information will be removed to ensure your privacy is protected. During the study, the researchers involved with the study, and authorised staff who work for, or with, the sponsor of the study, may require access to your

personal information and/or medical records to verify the data for this study and ensure that it is being conducted in accordance with UK law. All information will be treated in the strictest confidence during

Where this information could identify you, the information will be held securely with strict arrangements about who can access the information. The information will only be used for the purpose of health and care research, or to contact you about future opportunities to participate in research. It will not be used to make decisions about future services available to you, such as insurance. Where there is a risk that you can be identified your data will only be used in research that has been independently reviewed by an ethics committee.

What will happen to the results of the study?

The results of the study will be anonymous by the time they are in the public domain, and you will not be able to be identified from any of the data produced. When the results of this study are available they may be written into a reports for internal purposes and to inform the funder of the research, published in publically-accessible peer reviewed scientific journals and used for scientific presentations and conferences. We also plan to produce a summary of the results which will be easy for study participants to read and understand. With your consent we will send this summary to you.

Who is organising (sponsoring) and funding the study?

This study is sponsored by The University of Liverpool. The study is being conducted as part of a Doctoral Clinical Psychology project to help understand and treat the causes of chronic pain, and as such is as not externally funded.

Who has reviewed this study?

All research involving the NHS is reviewed by an independent group of people called a Research Ethics Committee (REC), to protect your interests.

Further information and contact details

Please feel free to contact a member of the study team listed below for any further information on tel. XXXX or email pain@liverpool.ac.uk

Christopher Knaggs

Dr Christopher Brown

For written correspondence:

Christopher Brown, 1.59c Eleanor Rathbone Building, Bedford Street South, Liverpool, United Kingdom, L69 7ZA.

immediate research team.

Your information could be used for research in any aspect of health or care, and could be combined with information about you from other sources held by researchers, the NHS or government.

Appendix H

NHS Research Ethics Committee Approval Letter

NHS
Health Research Authority
London - Bloomsbury Research Ethics Committee
HRA RES Centre Manchester
3rd Floor Barlow House
4 Minshull Street
Manchester
M1 3DZ

Telephone: 02071048285

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

15 January 2021

Dr Christopher Brown
Room 1.59c, 1st Floor, Eleanor Rathbone Building, University of Liverpool
Bedford Street South
Liverpool
L69 7ZA

Dear Dr Brown

Study title:	**Towards a neuropsychological profile of chronic pain: Development and validation of a novel measure of somatosensory working memory deficits in fibromyalgia.**
REC reference:	**20/PR/0545**
Protocol number:	**UoL001551 7809**
IRAS project ID:	**281672**

Thank you for your letter of 17 December 2020, responding to the Research Ethics Committee's (REC) 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-Chair, along with Lead and Second Reviewers, Miss Allyson Gray and Dr Michael Jacobs.

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, subject to the conditions specified below.

Good practice principles and responsibilities

The UK Policy Framework for Health and Social Care Research sets out principles of good practice in the management and conduct of health and social care research. It also outlines the responsibilities of individuals and organisations, including those related to the four elements of research transparency:

1. registering research studies
2. reporting results
3. informing participants
4. sharing study data and tissue

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations

Registration of Clinical Trials

All research should be registered in a publicly accessible database and we expect all researchers, research sponsors and others to meet this fundamental best practice standard.

It is a condition of the REC favourable opinion that **all clinical trials are registered** on a publicly accessible database within six weeks of recruiting the first research participant. For this purpose, 'clinical trials' are defined as the first four project categories in IRAS project filter question 2. Failure to register a clinical trial is a breach of these approval conditions, unless a deferral has been agreed by or on behalf of the Research Ethics Committee (see here for more information on requesting a deferral:
https://www.hra.nhs.uk/planning-and-improving-research/research-planning/research-registration-research-project-identifiers/

If you have not already included registration details in your IRAS application form, you should notify the REC of the registration details as soon as possible.

Further guidance on registration is available at:
https://www.hra.nhs.uk/planning-and-improving-research/research-planning/transparency-responsibilities/

Publication of Your Research Summary

We will publish your research summary for the above study on the research summaries section of our website, together with your contact details, no earlier than three months from the date of this favourable opinion letter.

Should you wish to provide a substitute contact point, make a request to defer, or require further information, please visit:
https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/

N.B. If your study is related to COVID-19 we will aim to publish your research summary within 3 days rather than three months.

During this public health emergency, it is vital that everyone can promptly identify all relevant research related to COVID-19 that is taking place globally. If you haven't already done so, please register your study on a public registry as soon as possible and provide the REC with the registration detail, which will be posted alongside other information relating to your project. We are also asking sponsors not to request deferral of publication of research summary for any projects relating to COVID-19. In addition, to facilitate finding and extracting studies related to COVID-19 from public databases, please enter the WHO official acronym for the coronavirus disease (COVID-19) in the full title of your study. Approved COVID-19 studies can be found at:
https://www.hra.nhs.uk/covid-19-research/approved-covid-19-research/

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

After ethical review: Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study, including early termination of the study
- Final report
- Reporting results

The latest guidance on these topics can be found at
https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/.

Ethical review of research sites

NHS/HSC sites

The favourable opinion applies to all NHS/HSC sites listed in the application subject to confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or management permission (in Scotland) being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS/HSC sites

I am pleased to confirm that the favourable opinion applies to any non-NHS/HSC sites listed in the application, subject to site management permission being obtained prior to the start of the study at the site.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Confirmation of any other Regulatory Approvals (e.g. CAG) and all correspondence [DClin RRC Approval Letter]	1	28 April 2020
Copies of materials calling attention of potential participants to the research [Advert Poster Healthy Controls]	2	02 December 2020
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [TWIMC Clinical Trials - 27.07.20]	1	15 September 2020
IRAS Application Form [IRAS_Form_18092020]		18 September 2020
Letter from sponsor [Letter(1).pdf]	1	21 August 2020
Non-validated questionnaire [Participant_Experiences AWM]	2	02 December 2020
Other [Demographic and Medication Questionnaire]	1	02 December 2020
Other [EOI form AWM]	2	02 December 2020
Other [Patient_invite_letter AWM Dec 2020]	2	02 December 2020
Other [Responses to REC Provisional Opinion]	1	17 December 2020
Participant consent form [Consent form AWM Experimental Group]	2	02 December 2020
Participant consent form [Consent form AWM Control Group]	2	02 December 2020
Participant information sheet (PIS) [PIS AWM Experimental Group]	2	02 December 2020
Participant information sheet (PIS) [PIS AWM Controls]	2	02 December 2020
Research protocol or project proposal [Pain Project Protocol]	2	02 December 2020
Summary CV for Chief Investigator (CI) [CV - Brown - 2 page]	1	07 September 2020
Summary CV for student [Chris Knaggs CV]	1	15 September 2020
Summary CV for supervisor (student research) [HannahTwiddy CV]	1	23 September 2020

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at:
https://www.hra.nhs.uk/planning-and-improving-research/learning/

IRAS project ID: 281672 Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely
p.p.

Reverend Jim Linthicum
Chair

Email:bloomsbury.rec@hra.nhs.uk

Enclosures: "After ethical review – guidance for researchers" *[SL-AR2]*

Copy to: Mr Alex Astor

Appendix I

Health Research Authority Approval Letter

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Dr Christopher Brown
Room 1.59c, 1st Floor, Eleanor Rathbone Building,
University of Liverpool
Bedford Street South
Liverpool
L69 7ZA

Email: approvals@hra.nhs.uk
HCRW.approvals@wales.nhs.uk

15 January 2021

Dear Dr Brown

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Towards a neuropsychological profile of chronic pain: Development and validation of a novel measure of somatosensory working memory deficits in fibromyalgia.**
IRAS project ID:	**281672**
Protocol number:	**UoL001551 7809**
REC reference:	**20/PR/0545**
Sponsor	**University of Liverpool**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?

HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report

(including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **281672**. Please quote this on all correspondence.

Yours sincerely,
Damilola Odunlami

Approvals Specialist

Email: approvals@hra.nhs.uk

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
Confirmation of any other Regulatory Approvals (e.g. CAG) and all correspondence [DClin RRC Approval Letter]	1	28 April 2020
Contract/Study Agreement template [mPICa]		
Copies of materials calling attention of potential participants to the research [Advert Poster Healthy Controls]	2	02 December 2020
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [TWIMC Clinical Trials - 27.07.20]	1	15 September 2020
IRAS Application Form [IRAS_Form_18092020]		18 September 2020
Letter from sponsor [Letter(1).pdf]	1	21 August 2020
Non-validated questionnaire [Participant_Experiences AWM]	2	02 December 2020
Other [Demographic and Medication Questionnaire]	1	02 December 2020
Other [EOI form AWM]	2	02 December 2020
Other [Patient_invite_letter AWM Dec 2020]	2	02 December 2020
Other [Responses to REC Provisional Opinion]	1	17 December 2020
Participant consent form [Consent form AWM Experimental Group]	2	02 December 2020
Participant consent form [Consent form AWM Control Group]	2	02 December 2020
Participant information sheet (PIS) [PIS AWM Controls]	2	02 December 2020
Participant information sheet (PIS) [PIS AWM Experimental Group]	2	02 December 2020
Research protocol or project proposal [Pain Project Protocol]	2	02 December 2020
Summary CV for Chief Investigator (CI) [CV - Brown - 2 page]	1	07 September 2020
Summary CV for student [Chris Knaggs CV]	1	15 September 2020
Summary CV for supervisor (student research) [HannahTwiddy CV]	1	23 September 2020

Appendix J

DClin RRC Outcome Letter

UNIVERSITY OF LIVERPOOL

D.Clin.Psychology Programme
Division of Clinical Psychology
Whelan Building, Quadrangle
Brownlow Hill
LIVERPOOL
L69 3GB

Tel: 0151 794 5530/5534/5877
Fax: 0151 794 5537
www.liv.ac.uk/dclinpsychol

Christopher Knaggs
Clinical Psychology Trainee
Doctorate in Clinical Psychology Programme
University of Liverpool
L69 3GB

28 April 2020

RE: Towards a neuropsychological profile of chronic pain: Development and validation of a novel measure of somatosensory working memory deficits in fibromyalgia.
Trainee: Christopher Knaggs
Supervisors: Christopher Brown and Hannah Twiddy

Dear Chris,

Thank you for your notification of amendment to your proposal submitted to the Chair of the D.Clin.Psychol. Research Review Committee.

Please take this Chairs Action decision as ***final*** approval from the committee.

You may now progress to the next stages of your research.

I wish you well with your research project.

Dr Ross White
Vice Chair D.Clin.Psychol. Research Review Committee

Appendix K

Medication and Demographic Questionnaire

Section 1. Demographic survey – all participants

PPT…
What is your **age** in years?
☐ years

What is your **gender**?
Male Female Other

Have you been experiencing **pain** that persists (is present on most days) or recurs (as repeating episodes) for the PAST THREE MONTHS or more?
Yes No

[If answer to the above is 'No', this is to make sure the healthy participants are not in pain while completing the study; participants who answer no to this question please move to Section 3]
Are you experiencing **pain** RIGHT NOW?
Yes No

Section 2. Medication checklist – participants with chronic pain only

What **treatments or medications** do you use REGULARLY (on most days) to relieve pain symptoms? Please select all that apply.

I do not use any treatments or medications for my pain	
Non-steroidal anti-inflammatory drugs (NSAIDs)	e.g. aspirin, ibuprofen, naproxen, etoricoxib
Paracetamol	
Mild opioids	e.g. codeine, low dose tramadol, meptazinol
Strong opioids	e.g. morphine, tapentadol, fentanyl, oxycodone, high dose tramadol
Antiepileptics	e.g. gabapentin, pregabalin
Antidepressants	e.g. amitriptyline, duloxetine, fluoxetine, nortriptyline, mirtazapine, certroline, diazepam
Steroids	e.g. prednisone, dexamethasone
Topical analgesics	e.g. lidocaine, capsaicin, rubefacients anti-inflammitory gel (volterol)
Psychological therapies	e.g. cognitive behavioural therapy, meditation, relaxation, biofeedback
Physical therapies	e.g. physiotherapy, occupational therapy, manual

1

therapy,

Electrotherapy	e.g. transcutaneous electrical nerve simulation (TENS), spinal cord stimulator
Complementary and alternative therapies	e.g. acupuncture, dietary supplements (vitamin D), q enzine q10, turmeric capsules, folic acid capsules herbal medicine
Other (please specify)	Blood thinners

What **treatments or medications** do you use 'AS NEEDED' (for 'breakthrough' pain/pain attacks/flare-ups) to relive pain symptoms? Please select all that apply.

I do not use any treatments or medications for my pain	
Non-steroidal anti-inflammatory drugs (NSAIDs)	e.g. aspirin, ibuprofen, naproxen, etoricoxib
Paracetamol	
Mild opioids	e.g. codeine, low dose tramadol, meptazinol
Strong opioids	e.g. morphine, tapentadol, fentanyl, oxycodone, high dose tramadol
Antiepileptics	e.g. gabapentin, pregabalin
Antidepressants	e.g. amitriptyline, duloxetine, fluoxetine, nortriptyline, mirtazapine
Steroids	e.g. prednisone, dexamethasone
Topical analgesics	e.g. lidocaine, capsaicin, rubefacients
Psychological therapies	e.g. cognitive behavioural therapy, **meditation**, relaxation, biofeedback mindfulness
Physical therapies	e.g. physiotherapy, acupuncture occupational therapy, manual therapy, exercise
Electrotherapy	e.g. transcutaneous electrical nerve simulation (TENS), spinal cord stimulator, pain gun
Complementary and alternative therapies	e.g. acupuncture, dietary supplements vit B12, herbal medicine,
Other (please specify)	

In the last 24 hours, how much relief have pain treatments or medications provided? Please select the one percentage that most shows how much **relief** you have received.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
No relief										Complete relief

Section 3. Fatigue Rating

Please rate the level of **fatigue** you feel RIGHT NOW.

0	1	2	3	4	5	6	7	8	9	10
iot tired, full of energy										Total exhaustion

Appendix L

Generalise Anxiety Disorder Assessment 7 (GAD7)

GAD-7 Anxiety

Over the last two weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid, as if something awful might happen	0	1	2	3

Column totals _____ + _____ + _____ + _____ =

Total score _____

If you checked any problems, how difficult have they made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all	Somewhat difficult	Very difficult	Extremely difficult
☐	☐	☐	☐

Source: Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD-PHQ). The PHQ was developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke, and colleagues. For research information, contact Dr. Spitzer at rls85@columbia.edu. PRIME-MD® is a trademark of Pfizer Inc.

Scoring GAD-7 Anxiety Severity

This is calculated by assigning scores of 0, 1, 2, and 3 to the response categories, respectively, of "not at all," "several days," "more than half the days," and "nearly every day."
GAD-7 total score for the seven items ranges from 0 to 21.

0–4: minimal anxiety

5–9: mild anxiety

10–14: moderate anxiety

15–21: severe anxiety

Appendix M

Patient Health Questionnaire

PATIENT HEALTH QUESTIONNAIRE (PHQ-9)

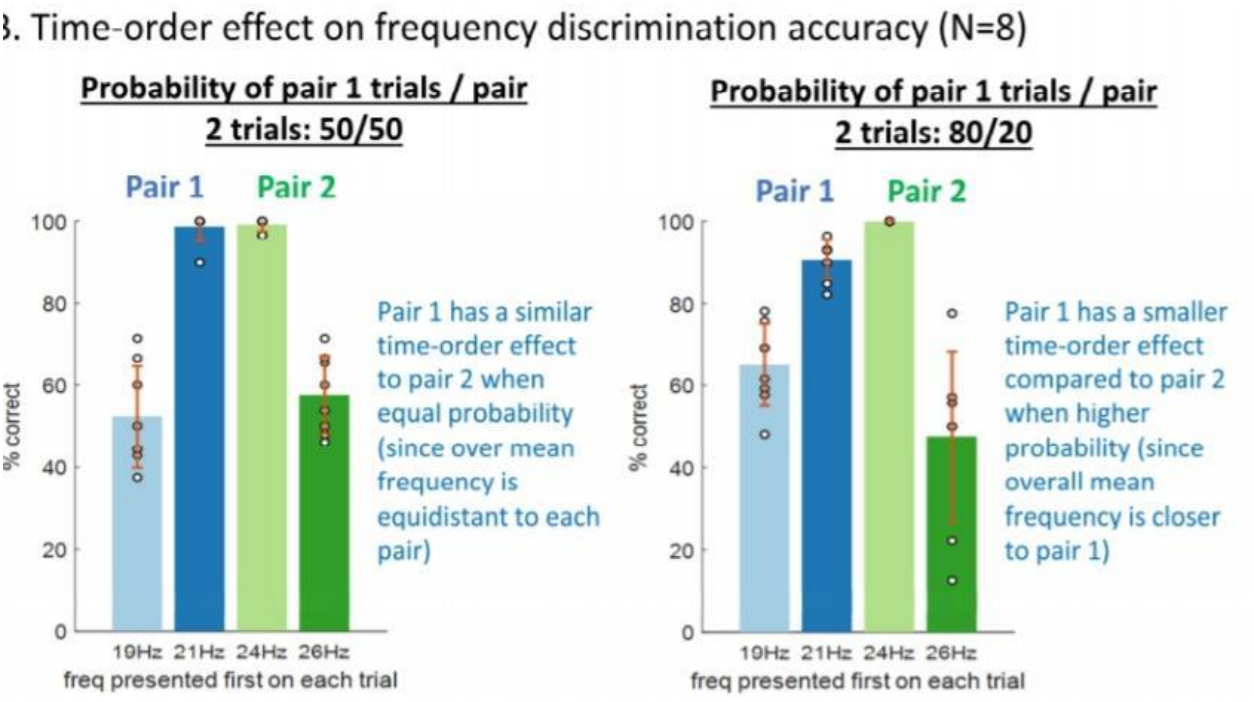
ID #: _____ **DATE:** _____

Over the last 2 weeks, how often have you been bothered by any of the following problems? (use "✓" to indicate your answer)	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite — being so figety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead, or of hurting yourself	0	1	2	3
	add columns		+	+
(Healthcare professional: For interpretation of TOTAL, please refer to accompanying scoring card).	TOTAL:			

10. If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?	Not difficult at all _____ Somewhat difficult _____ Very difficult _____ Extremely difficult _____

Appendix N

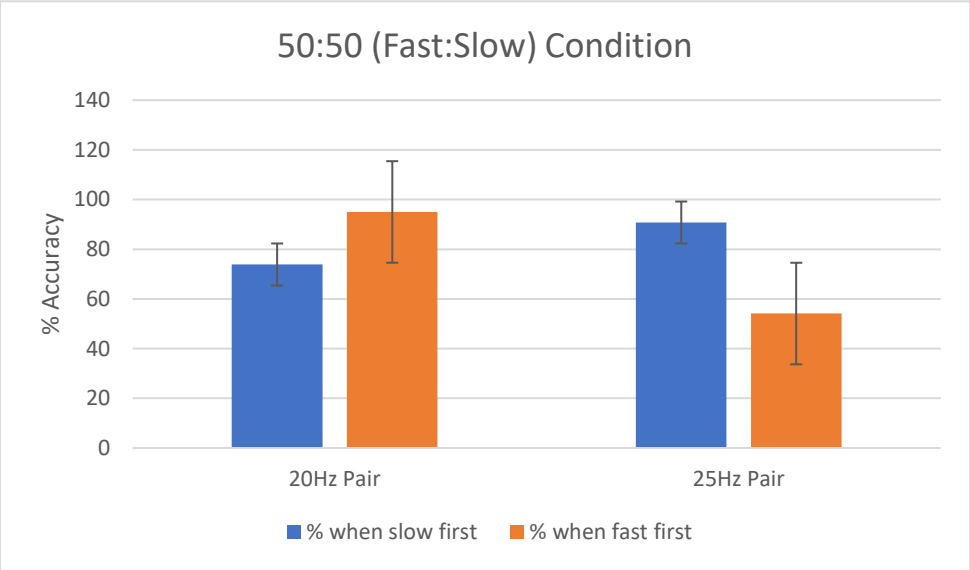
Pilot data from previous WM pilot study by researchers involved in the current study using somatosensory stimuli demonstrating variation in TOE between 50:50 and 80:20 blocks



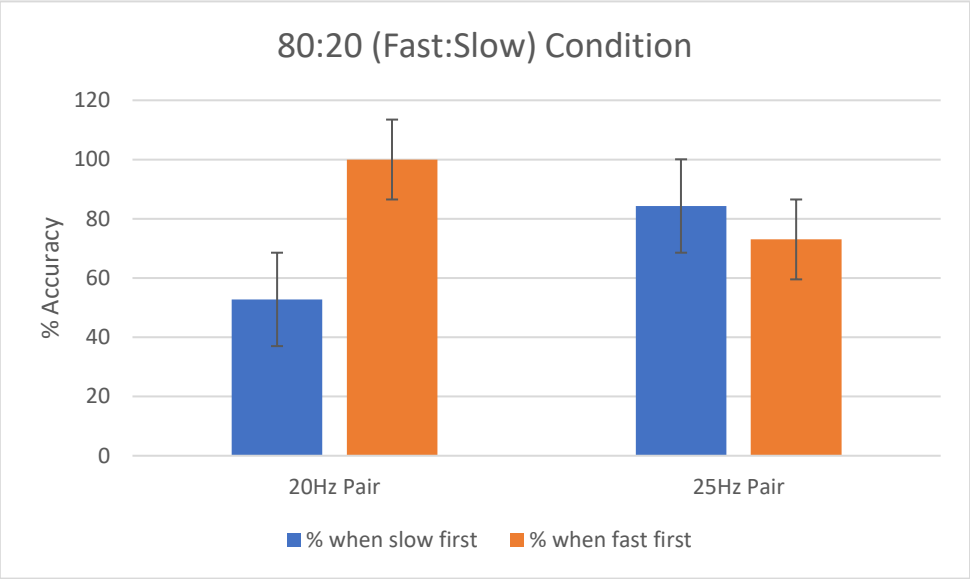
Appendix O

Pilot data conducted using healthy participants (n = 8) and the same procedures as the current study

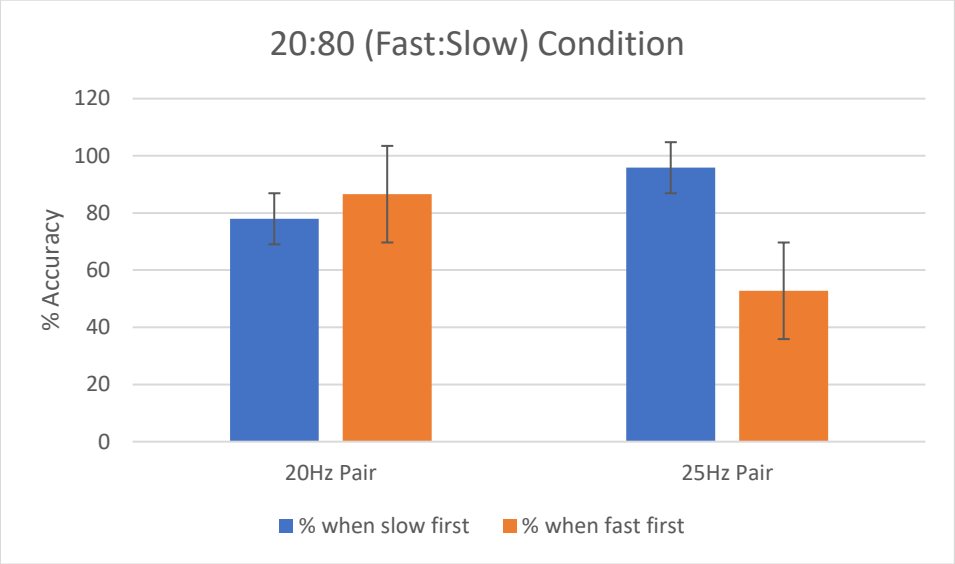
A



B



C



In A above it can be seen that in the 50:50 condition a classic TOE has been produced. Response accuracy in both pairs is highest for those trials where the frequency closest to the mean of all 4 frequencies in the condition is presented first. As hypothesised, in B and C it can be seen that the difference in response accuracy is less pronounced for those stimulus train pairs which are closest to the mean (20Hz and 25Hz respectively). Also as hypothesised, B and C the difference in response accuracy is more pronounced for those stimulus train pairs which are further away from the mean (25Hz and 20Hz respectively).