Involvement and engagement of seldom-heard communities in big data research

Thesis submitted in accordance with the requirements of the University of Liverpool
for the degree of Doctor in Philosophy

by

Piotr Teodorowski

August 2023

PAGE IS INTENTIONALLY LEFT BLANK

Table of Contents

Abstract 11
List of Tables 12
List of Figures 13
Declaration 14
Acknowledgements 15
Abbreviations 17
1. Thesis Introduction 18
  1.1. Research background 18
  1.2. Outline of thesis 18
  1.3. Terminology 23
    1.3.1. Involvement and engagement 23
    1.3.2. Who participates in the involvement? 24
    1.3.3. Seldom-heard communities 26
    1.3.4. Big data 27
  1.4. Conclusions 28
2. Public involvement 30
  2.1. Introduction 30
  2.2. History and policy context in the United Kingdom 30
    2.2.1. Beginning of the NHS 30
    2.2.2. Thatcherism 31
    2.2.3. New Labour 31
    2.2.4. One-nation conservatism 32
    2.2.5. Where is PPI now? 33
  2.3. The case for public involvement in health research 34
    2.3.1. Person-centred healthcare and research 34
    2.3.2. Moral and legal reasons for PPI 36
    2.3.3. Ideological arguments for PPI 37
    2.3.4. Researchers' views on PPI 38
  2.4. Involvement in practice 40
    2.4.1. What makes people enthusiastic, interested and committed to PPI? 40

2.4.2. The importance of involving seldom-heard communities 41
2.5. Involving seldom-heard communities 42
2.5.1. Inclusive environments 42
2.5.2. Capacity building 44
2.6. Conclusions 46
3. Big data 47
3.1. Introduction 47
3.2. Public views towards big data research 47
3.3. What are public concerns around big data? 49
3.3.1. Lack of consent 49
3.3.2. Confidentiality 50
3.3.3. Data misuse 50
3.4. Public controversies around big data research 51
3.4.1. Care.data 52
3.4.2. Google Deep Minds 52
3.4.3. General Practice Data for Planning and Research 53
3.5. Social licence for big data research 55
3.6. Conclusions 58
4. Review protocol 59
4.1. Overview 59
4.2. Abstract 60
4.3. Introduction 61
4.4. Review objectives 64
4.5. Methods and analysis 64
4.5.1. Design 64
4.5.2. Stage 1: Scoping review 65
4.5.3. Stage 2: Systematic review 69
4.6. Patient and public involvement 70
4.7. Limitations 70
4.8. Ethics and dissemination 70
4.9. Conclusions 70

5. Review findings 72
5.1. Introduction 72
5.2. Methods 72
5.2.1. Searches 72
5.2.2. Screening 72
5.2.3. Data extraction and analysis 72
5.3. Stage 1: Scoping Review 73
5.3.1. Study Characteristics 74
5.3.2. Activities 80
5.3.3. Outcomes 83
5.3.4. References to involvement or engagement guidance 83
5.3.5. References to legal documents 84
5.3.6. Challenges to involvement and engagement 84
5.3.7. Facilitators of involvement and engagement 86
5.4. Stage 2: Systematic Review 88
5.5. Discussion 88
5.6. Revised model 90
5.7. Limitations 91
5.8. Conclusion 91
6. Methodology 93
6.1. Introduction 93
6.2. Philosophical Underpinnings 93
6.2.1. Social constructionism as an approach to social science research 93
6.2.2. Features of social constructionism 95
6.3. Qualitative research 97
6.3.1. Methods 100
6.3.1.1. Qualitative interviewing 100
6.3.1.2. Sampling strategy 105
6.3.2. Recruitment 107
6.3.3. Interview procedure 109

6.3.4. Twitter data 113

6.3.5. Data saturation 114

6.4. Reflexivity & positionality 116
6.4.1. Personal and functional reflexivity 117

6.4.2. Disciplinary reflexivity 119

6.5. Ethical considerations 120
6.5.1. Ethical approval 120

6.5.2. Voluntary participation & informed consent 121

6.5.3. Confidentiality 122

6.5.4. Potential benefits & risks 122

6.5.5. Data storage 123

6.6. Conclusions 123
7. Data analysis 125
7.1. Introduction 125
7.2. Reflexive thematic analysis 125
7.3. Conducting reflexive thematic analysis 127
7.3.1. Familiarizing oneself with the data 127

7.3.2. Generating codes 129

7.3.3. Constructing themes, 131

7.3.4. Reviewing potential themes, 133

7.3.5. Defining and naming themes, 135

7.3.6. Producing the report 136

7.4. Orienting to data 137
7.4.1. Big theory 137

7.4.2. Small theories 138

7.5. Limitations of and misconceptions around this approach to data analysis 140
7.6. Reflections 141
7.7. Conclusions 142
8. Public involvement in the thesis 143
8.1. Introduction 143

8.2. Setting up and recruitment through the Public Advisors Forum 143
8.3. The literature review 146
8.4. Interviews with Polish and South Asian communities 147
8.5. Interviews with Researchers 149
8.6. Challenges of public involvement in doctoral research 149
  8.6.1. Power relationships 149
  8.6.2. Positionality of public contributors 150
  8.6.3. Researching within the confines of the field 151
  8.6.4. Working with public contributors as a learning process 151
8.7. Impact of the Covid-19 pandemic 152
8.8. Conclusions 152
9. Twitter study 154
9.1. Linking the literature review with the paper 154
9.2. Abstract 155
9.3. Introduction 156
9.4. Models of public engagement 158
  9.4.1. Deficit model 158
  9.4.2. Dialogue model 159
  9.4.3. Participatory model 159
9.5. Research questions 159
9.6. Methods 160
  9.6.1. Data collection 160
  9.6.2. Analysis 161
  9.6.3. Ethics 162
9.7. Results 162
  9.7.1. Descriptive statistics 162
  9.7.2. Attitudes 163
  9.7.3. Thematic analysis 165
9.8. Discussion 169
9.9. Limitations 172
9.10. Conclusions 172

10. Researchers' study 174
10.1. Linking the literature review with the paper 174
10.2. Abstract 175
10.3. Introduction 176
10.3.1. Big data 176
10.3.2. Seldom-heard communities 177
10.4. Research aim 178
10.5. Methods 178
10.5.1. Theoretical position 178
10.5.2. Participants and data collection 178
10.5.3. Data analysis 179
10.5.4. Ethics statement 180
10.6. Findings 180
10.6.1. Abstraction and complexity of big data 180
10.6.2. One size does not fit all 182
10.6.3. Working in partnership 184
10.6.4. Empowering the public contribution 186
10.7. Discussion 188
10.8. Study limitations 190
10.9. Conclusion 190
11. Polish and South Asian communities' study 192
11.1. Linking the literature review with the paper 192
11.2. Abstract 193
11.3. Introduction 194
11.3.1. Theoretical underpinning 196
11.3.2. Research aims 197
11.4. Methods 197
11.4.1. Participants 197
11.4.2. Procedure 198
11.4.3. Public contributors 199
11.4.4. Data analysis 200

11.4.5. Ethics 200
11.5. Results 200
11.5.1. Happy to reuse data 201
11.5.2. Capability 203
11.5.3. Motivation 205
11.5.4. Opportunity 209
11.6. Discussion 211
11.7. Strengths and limitations 213
11.8. Conclusion 213
12. Discussion 215
12.1. Introduction 215
12.2. Synopsis of findings 215
12.2.1. RQ1: How can social media be used to facilitate the engagement process? 215
12.2.2. RQ2: What are the experiences of researchers when involving and engaging seldom-heard communities in big data research? 218
12.2.3. RQ3: How to involve seldom-heard communities that previously have not been involved in big data research? 220
12.2.4. How do these findings contribute to the research aims? 222
12.3. How do these findings add to the literature on public involvement and engagement in big data research? 223
12.3.1. System logic model of public involvement and engagement in big data research 224
12.3.2. Adapting the COM-B model to public involvement 226
12.3.3. Recommendations for involving and engaging seldom-heard communities in big data research 227
12.4. Strengths and limitations of this doctoral research 231
12.5. Future directions for public involvement and engagement in big data research 234
12.6. Future of public involvement and engagement in big data research 235
12.7. Concluding remarks 238
References 239
Appendices 263

Appendix 1: Data extraction form 263
Appendix 2: Topic guide for interviews with Polish & South Asians 264
Appendix 3: Topic guide for interviews with researchers 266
Appendix 4: Example of participant information sheet 268
Appendix 5: Approval letter from ethics committee for Twitter study 271
Appendix 6: Approval letter from ethics committee for interviews with seldom-heard communities and researchers 273
Appendix 7: Public contributors call shared with the Public Advisors Forum 276
Appendix 8: Visual minutes from public involvement sessions 277
Appendix 9: Further examples of involvement and engagement in big data research from researcher’s perspectives. 278
Appendix 10 Standards for Reporting Qualitative Research (SRQR)* 280

Abstract

Involvement and engagement of seldom-heard communities in big data research

Piotr Teodorowski

Ongoing digitalisation means researchers can reuse previously collected data for new purposes such as research. There have been public controversies in the United Kingdom around big data research. One way to address public concerns could be greater public involvement and engagement. However, little research has been conducted to see how to involve and engage members of the public in big data research, especially those from seldom-heard communities.

Therefore, my research explored how to involve and engage members of the public, especially seldom-heard communities, in big data research. Initially, I conducted the scoping review to map the current evidence around public involvement and engagement in big data research. Then, I conducted three primary studies to address the research aim. First, the Twitter study examined the Data Saves Lives campaign and how the public could be engaged on social media. Second, I interviewed researchers (n=12) with experience in involving and engaging seldom-heard communities in big data research. Third, I interviewed Polish (n=20) and South Asian (n=20) participants who were not previously public contributors to understand how they could be involved in big data research. The data was analysed using reflexive thematic analysis. Two public contributors were involved in the research by supporting design, ethics application, recruitment, analysis and dissemination.

This thesis contributes to existing public involvement and engagement literature by drawing on more examples of how to involve and engage members of the public in big data research. The social media discourse around the Data Saves Lives campaign was mostly positive towards big data research, but the results indicate that organisations and professions drive the discussion; thus, the campaign might have rarely engaged directly with the public. Researchers' study offers an in-depth exploration of facilitators and barriers in involving seldom-heard communities. Four themes illustrate the participants' experiences: (1) abstraction and complexity of big data, (2) one size does not fit all, (3) working in partnership, and (4) empowering the public contribution. The interviews with Polish and South Asian communities present their views towards big data research and when their data could be reused. I mapped the themes around involvement from these interviews under the capability-opportunity-motivation-behaviour (COM-B) model to see how different factors influence participants' willingness to become public contributors.

This thesis offered a new body of evidence on how to involve and engage members of the public, especially coming those from seldom-heard communities, in big data research. The findings suggest that social media can potentially engage the wider public around big data research, but currently, Twitter does not utilise its full potential to achieve it. I showed how to adapt the COM-B model to public involvement. Based on the findings from the scoping review, I created the logic model of public involvement and engagement to map the aspects of the process. Last, I reflected on how public involvement in big data research might develop in the future and recommended further studies in the field.

List of Tables

Table 1 Search strategy 67
Table 2 Summary of included papers in the scoping review 80
Table 3 Summary of data saturation for each study 116
Table 4 Example of coding interview with a South Asian participant 131
Table 5 Example of theme narrative undertaken during reflexive thematic analysis 136
Table 6 Questions guiding mapping of findings from thematic analysis under the COM-B model 139
Table 7 Summary of the public involvement, training offered and impact in the thesis. 146
Table 8 Locations of Twitter users utilising #DataSavesLives 162
Table 9 The list of the ten most active Twitter accounts using #DataSavesLives 163
Table 10 Themes and subthemes derived from the #DataSavesLives debate on Twitter through reflexive thematic analysis. 165
Table 11 Key recommendations on involving and engaging seldom-heard communities around big data research 188
Table 12 The summary of the thesis's contributions to the field. 231
Table 13 Principles and standards for public involvement and engagement 236

List of Figures

Figure 1 System logic model of public involvement and engagement in big data research...63
Figure 2 Systematic map of the review process...65
Figure 3 PRISMA flow chart as adopted by Moher *et al.* (2009) and Moher *et al.* (2015)...74
Figure 4 The updated system logic model of public involvement and engagement in big data research. Asterix * is used to record new aspects of the model based on the review. ...90
Figure 5 Initial thematic mind map for the Twitter study...133
Figure 6 Revised thematic map for the Twitter study...134
Figure 7 Process of cleaning data sets for qualitative analysis...161
Figure 8 Key connectors and relationships between themes and subthemes. ...165
Figure 9 COM-B model as developed by Michie and colleagues...197
Figure 10 Themes mapped under COM-B model...201
Figure 11 The final system logic model of public involvement and engagement in big data research. * new aspects of the model based on this thesis. ...225

Declaration

This thesis is the result of my own work. No portion of the thesis has been submitted either partially or otherwise in support of any other degree or qualification at this or any other institution.

I am the first author on all peer-review journal publication presented as part of this thesis. In the introduction, I include the authorship contributions for each paper.

Acknowledgements

First, I would like to acknowledge and thank my supervisors, Lucy Frith, Sarah Rodgers and Kate Fleming, for their guidance, support and encouragement. I am incredibly grateful for their expertise throughout my studies.

Thank you to all participants and public contributors for taking part in my research. My special thanks go to public contributors involved in my studies over all three years, Naheed Tahir and Saiqa Ahmed.

I would like to acknowledge all the professional support that I received during this research: Estefania Cordero (European Patients Forum), Gözde Susuzlu Briggs (European Patients Forum), Zoe Gibbs-Monaghan (University of Liverpool), Ruaraidh Hill (University of Liverpool), Colette Miller (ARC NWC) and Selina Wallis (ARC NWC). I would also like to thank my fellow PhD student Elisa Jones for her supportive friendship during my studies.

I am most appreciative to my partner, Nico. There are no words to express my gratitude for her support, tolerance and believing in me. Further thanks go to Mickey and Donald, who ensured that I had regular writing breaks.

Last, I would like to acknowledge the support of the National Institute for Health and Care Research Applied Research Collaboration North West Coast (NIHR ARC NWC) for funding my doctoral research and making this PhD possible.

In loving memory of Donald.

8th December 2023

Abbreviations

ARC NWC Applied Research Collaboration North West Coast

COM-B model Capability-Opportunity-Motivation-Behaviour model

GPDPR General Practice Data for Planning and Research

HRA Health Research Authority

NHS National Health Service

NIHR National Institute for Health and Care Research

PPI Patient & Public Involvement

PPIE Patient & Public Involvement and Engagement

UK United Kingdom

1. Thesis Introduction

This PhD thesis presents research exploring how to involve and engage the public (especially seldom-heard communities) in big data research. In this chapter, I provide an overview of the research background. Then, I outline the thesis chapters and peer-reviewed published papers that have been incorporated into the thesis. Finally, I discuss key terms that are used throughout this work.

1.1. Research background

The stimulus for this research has been the ongoing digitalisation and the growth of big data research that uses the linkage of large, multi-source health data sets to offer new research applications. It could be argued that the success of big data research depends on public support. For example, a previous big data project called the care.data lacked public support, which eventually led to its shutdown (Hays and Daker-White, 2015). Public involvement and engagement could be one of the ways of establishing public support. All communities need to be involved and engaged; if big data research aims to examine and tackle health inequalities, then a special focus should be on seldom-heard groups as these are often less involved in research, but experience more health inequalities than the general population. Not involving all sections of society might lead to their views being excluded. However, there has been limited understanding of how to involve and engage seldom-heard communities in big data research.

Therefore, the overall aim of this thesis was to explore how to involve and engage the public (especially seldom-heard communities) in big data research. This was achieved by addressing the following research questions (RQ):

- RQ1: How can social media facilitate the engagement process?
- RQ2: What are the experiences of researchers when involving and engaging seldom-heard communities in big data research?
- RQ3: How to involve seldom-heard communities that previously have not been involved in big data research?

1.2. Outline of thesis

This thesis consists of peer-review papers[1] and traditional chapters. The following outline describes each contribution:

---

[1] These included peer-reviewed papers published in academic journals. The text in this thesis can slightly differ from the final version published online because here accepted authors manuscripts are incorporated.

1. Thesis overview

In this introductory chapter, I provide the research overview and briefly discuss the terms used throughout this thesis.

2. Public involvement and engagement

This is the first of four-literature review chapters and focuses on public involvement in healthcare and research in the UK. The chapter provides a cohesive summary of existing knowledge in the field and highlights gaps in public involvement literature. It starts with tracking the historical development of public involvement in the UK from the beginning of the NHS. It explores the current role of public involvement in research and healthcare services. Then, I present arguments for involvement, looking at principles of person-centred healthcare, moral, legal and ideological arguments, and discuss researchers' perspectives towards public involvement. This follows with an overview of what is known about how public involvement looks in practice and how the previous research suggests involving seldom-heard communities.

3. Big data research

This second literature review chapter focuses on big data research and thus aims to provide a context where public involvement and engagement occur for this thesis. I look at public awareness and views towards big data research. Then, I explore three public controversies in the UK around big data projects and see how the public response towards them influenced big data initiatives. Finally, I present the concept of social licence for big data research and examine what the literature says about the role of public involvement and engagement in ensuring social licence.

4. Review methodology

This chapter aims to bring together topics from two previous chapters by looking at public involvement and engagement in big data research. It is a published paper outlining the methodology for a two-stage review exploring public involvement and engagement in big data research. In the first stage, I conducted the scoping review exploring public involvement and engagement in big data research. In the second stage, out of the papers included in the scoping review, I extracted qualitative and quantitative projects that discussed the delivery and effectiveness of public involvement and engagement projects.

Teodorowski, P., Jones, E., Tahir, N., Ahmed, S. and Frith, L. (2021) 'Public involvement and engagement in big data research: protocol for a scoping review and a systematic review of delivery and effectiveness of strategies for involvement and engagement', *BMJ Open,* 11(8), pp. e050167.

Author contributions: PT developed the study design, drafted the protocol and conducted initial searches with the assistance of the librarian. EJ, NT, SA, LF contributed to drafting and editing. All authors have read and approved the final manuscript.

5. Literature Review

This chapter presents the findings from the review that was produced using the methodology presented in the previous chapter. It explores how public involvement and engagement are conducted in big data research and what is already known about this process. Using these findings, I identified gaps in public involvement and engagement in big data research literature that were explored in primary studies in this thesis. These gaps included a lack of primary research in three areas. First, despite the growth of social media, there has been limited understanding of how it could facilitate public engagement in big data research. Second, little was known about the experiences of big data researchers who involve and engage seldom-heard communities. Third, the perspectives and expectations of seldom-heard communities were underexplored, and studies did not look into the views of people who had not been previously involved.

6. Methodology

This chapter discusses theoretical underpinnings, methodology, and methods applied in this thesis. I examine the key features of social constructionism and how it influenced my decisions about research methods and approaches to data. Second, I discuss the methods used in this thesis by examining how I conducted qualitative interviews and collected Twitter data. Third, I reflect on my positionality and how it informed my research. After that, I cover ethical considerations for my research and show how I followed ethics standards and received ethical approvals from the University ethics committee.

7. Data analysis

Reflexive thematic analysis was used to analyse the data collected for this thesis. The chapter offers a brief justification for the choice of analytical method and reflects why this was the most appropriate approach to analyse my data. Second, I discuss each step of the reflexive thematic analysis in detail. This includes examples from primary studies. Third, I reflect on my orientation to data and how theoretical lenses influenced the analysis. Thereafter, I discuss the limitations of the reflexive thematic analysis and what common misconceptions exist around it. Finally, I conclude the chapter with my own reflection on this analytic approach.

8. Public involvement in the thesis

This chapter discusses how two public contributors have been involved throughout the PhD research and what impact it had on my work. First, I explain where and how I recruited them. Second, I outline their involvement throughout different projects that are a part of this thesis. This includes the impact of their involvement and how I supported them throughout this process. Thereafter, I reflect on the challenge of working with public contributors in doctoral research. Finally, the chapter concludes the discussion by recognising how Covid-19 impacted public involvement in my doctoral journey.

9. Twitter Study

This chapter is the first one presenting the findings from a primary study and is the second peer-reviewed published paper. It reports the results from the study exploring the Data Saves Lives campaign on Twitter. I collected tweets published between 19th April and 15th July 2021 that had the hashtag DataSaveLives. Thereafter, I coded the dataset twice. First, I assigned each tweet a positive, neutral or negative attitude towards the campaign. Second, using thematic analysis, I inductively coded data. Thereafter, I introduced theoretical lenses to the analysis by mapping themes from analysis under models of public engagement. These were deficit, dialogue and participatory. This allowed me to understand further how the campaign was utilised on Twitter.

Teodorowski, P., Rodgers, S. E., Fleming, K. and Frith, L. (2022) 'Use of the Hashtag# DataSavesLives on Twitter: Exploratory and Thematic Analysis', Journal of medical Internet research, 24(11), pp. e38232.

All authors contributed to the study design. PT drafted the first draft of this paper, and SER, KF, and LF contributed to drafting and editing. All authors have read and approved the final manuscript.

10. Interviews with researchers

The third published paper reports the findings from interviews with researchers who involved and engaged seldom-heard communities in big data research. I conducted interviews with twelve researchers who had experience involving or engaging seldom-heard communities in big data research. Then, I conducted the reflexive thematic analysis to analyse participants' experiences. These were reported under four themes. This study offered a better understanding of how to involve and engage the public in big data research.

Teodorowski, P., Rodgers, S. E., Fleming, K., Tahir, N., Ahmed, S. and Frith, L. (2023) ''To me, it's ones and zeros, but in reality that one is death': A qualitative study exploring researchers' experience of involving and engaging seldom-heard communities in big data research', *Health Expectations,* 26, pp: 882-91.

All authors contributed to the study design. PT drafted the first draft of this paper, and SER, KF, NT, SA and LF contributed to drafting and editing. All authors have read and approved the final manuscript.

11. Interviews with Polish and South Asian communities

The last published peer-reviewed paper explores how to involve seldom-heard groups in big data research by focusing on Polish and South Asian communities. In this project, I interviewed Polish (n=20) and South Asian (n=20) participants living in the United Kingdom who have not been previously involved as public contributors. Data were analysed using reflexive thematic analysis and then themes were mapped under the capability-opportunity-motivation-behaviour (COM-B) model (Michie, Van Stralen and West, 2011). Using the COM-B model allowed me to understand what could influence participants' willingness to become involved in research as public contributors.

Teodorowski, P., Rodgers, S. E., Fleming, K., Tahir, N., Ahmed, S. and Frith, L. (2023) Exploring how to improve the involvement of Polish and South Asian communities around big data research. A qualitative study using COM-B model., International Journal of Population Data Science, 8(1).

All authors contributed to the study design. PT drafted the first draft of this paper, and SER, KF, NT, SA and LF contributed to drafting and editing. All authors have read and approved the final manuscript.

12. Discussion

This chapter brings together findings from published papers. First, I locate them jointly in the literature around public involvement in big data research. Second, I discuss the strengths and limitations of this thesis. Thereafter, based on my thesis, I offer recommendations for researchers to involve and engage seldom-heard communities around big data research. Finally, I make some concluding remarks.

13. Appendices

The appendices consist of additional files that provide further information on the data analysis and include the ethics approvals.

1.3. Terminology

This section provides an overview of the terms used in the thesis. These include involvement and engagement, how to name people included in these activities, seldom-heard communities, and big data research. Each of them has been used inconsistently in academia and I clarify how I used them in this thesis.

1.3.1. Involvement and engagement

The Oxford English Dictionary defines involvement as "*the action or process of involving; the fact of being involved (…) entangled, or engaged; engagement*"(OED, 1990). Involvement, engagement, and participation are used interchangeably in the literature but do not necessarily have the same meaning (Mockford *et al.*, 2012; Islam and Small, 2020). This makes research and discussion about public involvement challenging, for example, identifying papers for reviews (Dawson *et al.*, 2018; Harrison *et al.*, 2019b; Lalani *et al.*, 2019). The most popular definition of involvement in the UK is by NIHR (2021a) (previously developed by INVOLVE (2020)[2]). It defines involvement as "*research being carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them*." This definition excludes engagement activities, promoting research or just building trust with the public. Therefore, NIHR defines engagement as activities "*where information and knowledge about research is provided and disseminated.*" Both involvement and engagement differ from participation which refers only to people who participate in the research (NIHR, 2021a). Engagement and involvement can be interrelated. Engagement could occur before involvement (to build links with the public) or after research (to disseminate its findings).

There is no one type of involvement. The Arnstein (1969) typology of involvement locates involvement stages on a ladder based on the amount of power given to the public. It identifies from the bottom (lowest extent of people influence) to the top (highest extent of people influence) the following steps: therapy manipulation, non-participation, informing, consultation, placation, partnership, delegation, and full citizen control. The author herself called the ladder 'provocative' (p216). Wilcox (1994) and Hart (1997) also utilise the ladder metaphor but suggest different wording. The former simplifies the ladder by suggesting the following steps:

---

[2] INVOLVE was established in 1996 by NIHR to support development of public involvement in health research. It has later ceased to exist and was incorporated into NIHR. Hence, when relevant I might cite any of these organisations.

information, consultation, deciding together, acting together and supporting initiatives. The latter presents specific steps for children and young people's participation: manipulation, decoration, tokenism, assigned but informed, consulted & informed, adult, initiated shared decisions with children, child-initiated & directed, child-initiated shared decisions with adults. Beresford (2012) criticises ladders as a too simplistic way of explaining the complex process. I agree with Beresford but also recognise that the metaphor can assist researchers in moving up the ladder to achieve better public involvement. Still, the metaphor's misleading part is that researchers should always start from the bottom and, if needed, climb up. Each research project or activity might require a different level of involvement, but the researcher should start from the top level of involvement and if they recognise that this is not feasible, move down.

Other terms can be confused with public involvement and engagement. McGrath (2016) points out that it can be a '*very inconvenient truth*' that involvement does not equal co-production. Key characteristics of co-production are:

> '*recognising people as assets, building on people's existing capabilities, promoting mutuality and reciprocity, developing peer support networks, breaking down barriers between professionals and recipients, facilitating rather than delivering*' (Boyle *et al.*, 2010, pp. 3-4)

Co-production takes involvement a step further. The public articulates their views and becomes a part of the research. Thus, co-production would be located above any of the discussed ladders of participation. However, co-production also remains an evolving concept. Masterson *et al.* (2022) identify in their review sixty commonly used definitions of co-production in health and social care research. The second term overlaps with both involvement and co-production is participatory research- an umbrella description of approaches arguing for co-ownership and co-decision between the public and researchers (Macaulay, 2016). It takes the power relationship to the next level and thus on the participation ladder it could also be located on the top.

Despite broad use, involvement and engagement remain evolving concepts in academia. In my work, I follow NIHR definitions. Two abbreviations are used throughout this thesis, PPI as patient and public involvement and PPIE as patient and public involvement and engagement. Who these 'patients and public' are is discussed next.

1.3.2. Who participates in the involvement?

Academics and practitioners use various terminology to describe involvement participants, and these have distinct origins in consumerism and democratic approaches (Beresford, 2003)[3]. These include patients, service users, consumers, the public, stakeholders and public contributors. Now, each meaning shall be examined, and I justify the choice of wording used in this thesis.

Patients, service users & consumers

Patients are '*experts by experience*' (NHS England, 2017, p. 6). Because they use healthcare services, they are also often called service users. Service users are broader than patients, as patients would refer only to health services, and service users, could also include social care. In government discourse, service users, patients, clients, users, and users of service have been utilised in the past (Hui and Stickley, 2007).

The usage of the word patients has received some criticism because it signifies the power imbalance between patients and health professionals (Neuberger and Tallis, 1999). Professionals decide on priorities and topics of discussion and have control over the relationship (Sharma, 2018). Thus, some would prefer the term service users (Beresford, 2012). The critics of the term patients claim that it suggests that medics know and understand sickness, so patients' roles are to do what medical professionals tell them to do. This contradicts what PPI aims to achieve in shifting the balance of power to more shared decision-making. The alternatives, such as consumer or client, suggest that healthcare is more about receiving service rather than an equal relationship (Neuberger and Tallis, 1999). However, the meaning of the consumer is not used consistently. Cochrane (2017, p. 2) practises the term but gives more power to those who get involved as their role can

> "*vary in extent and depth, from reviews which are guided by the contributions of consumers, to those where consumers control aspects of the review in which their lived experiences give them particular expertise (co-production).*"

However, patients would not perceive themselves as NHS consumers (Hogg, 2009). It might sound unnatural, as in the UK context, healthcare is publicly financed. Another alternative is 'citizens', but this would exclude migrant communities (Hogg, 2009).

The systematic review (Costa *et al.*, 2019) explored alternatives and found that patients remained the preferable term for those using healthcare. One of the included studies in the review was of Simmons *et al.* (2010), who found that patients is the preferred term through a questionnaire with over 300 healthcare professionals in mental health services. It is worth

---

[3] For more discussion on consumerism and democratic approaches in public involvement, please see literature review chapter at page 37

noticing the differences between some professions, as health professionals (but not social workers) dislike the term ‘service user’.

Public

The public is a broader term than patients or service users. It consists of

> “*patients, potential patients, carers and people who use health and social care services as well as people from organisations that represent people who use services*” (INVOLVE, 2020).

Often the terms public and patients have been used together to stress the inclusion of service users, potential users and their families and carers (NHS England, 2017).

Stakeholders

Stakeholder is the broadest term. Grotz, Ledgard and Poland (2020, p. 47) suggest that

> “*[stakeholder] can mean anything from the patient and the public involved in a research project; through everyone directly involved in the work of a project; to everyone who has a link to the project, including service providers and users, commissioners, research funders and many others*”.

Thus, Grotz and colleagues also argue that stakeholders can be used later as an excuse to avoid ordinary patients and the public, as involving other groups (including professionals) could be considered adequate.

Public contributors

Public contributors (alongside public advisors or public partners) are terms used to describe people directly involved in PPI activities.

In this section, I showed that diverse terms are used to describe people involved in PPI. Throughout this thesis, I use two main cluster terms. The terms public or patients are used throughout this thesis and mean the same thing, namely referring to lay members of the public. The term public contributor describes members of the public that participate in PPI.

1.3.3. Seldom-heard communities

The term seldom-heard communities has been evolving over the years and used alongside others such as hard to reach, seldom listened to, peripheral voices, marginalised, under-served, socially excluded and hidden groups. INVOLVE (2012a) defines them as "*people and groups often overlooked or ignored by mainstream society*". Freimuth and Mettger (1990) describe them (in the context of health campaigns) as "*persons who have a low socioeconomic* *status, members of ethnic minorities, and persons who have a low level of literacy*." The common characteristics is that policymakers, researchers and healthcare professionals miss them (Morrow, Boaz and Sally, 2011). There is a risk that those experiencing health inequalities might also be groups that get involved in PPI less than the general public (Dawson *et al.*, 2018). Ignoring them could make research findings ungeneralisable to all parts of society and miss nuances of opinions specific to those groups (Bonevski *et al.*, 2014).

However, there is no comprehensive list of who these groups are. It is impossible to name all potential seldom-heard groups. It can be misleading and counterproductive as people can be members of one group but do not share the same experiences. One community can be a seldom-heard group in one circumstance but not the other, so researchers should reflect on each study context (NIHR, 2020). For example, young people can be a seldom-heard community in a study looking at a local health service but a majority group in a project around social media. To understand which groups tend to be more left out, Beresford (2013) identifies key exclusion criteria. These include equality issues (gender, ethnicity, age etc.), where people live (e.g. homelessness, traveller communities), communication issues (e.g. deaf people or not native speakers), the nature of impairment (e.g. people living with dementia) and unwanted voices (public contributors that might be confident to raise comments unwelcomed by researchers). More recently, NIHR (2020) grouped potential seldom-heard communities (but not as an exhaustive list) within clusters of demographic factors (age, sex, ethnicity or education), social and economic factors, health status and disease-specific factors. This shows that researchers have a broad scope to reflect on which communities might be seldom-heard for their research project.

In this thesis, I perceive seldom-heard communities as those who participate less in PPIE than the general population. This could be caused by researchers' or members of the public inability to be involved or engaged. Although this remains context-specific- in this PhD, it is around big data research that I define in the next section.

1.3.4. Big data

The term big data has been first used by Cox and Ellsworth (1997) in their conference paper. Since then, it has been developing through the fast-paced advance of digital technology. Big data is still an evolving field (Aitken *et al.*, 2019), and disagreements remain on what the term exactly stands for in health research (Mehta and Pandit, 2018). However, a systematic review attempted to determine the scope of big data in health and included five key characteristics (Mehta and Pandit, 2018). These are called '5V' and stand for volume, velocity, variety, veracity and value. The volume is about the quantity of data as this is expected to keep growing. The variety means that it can come from both structured and unstructured sources.

The velocity refers to the speed with which data can transpire or even be collected in real time. The veracity is around the quality and accuracy of data. The value refers to the benefit to policymakers (or healthcare professionals) to make a meaningful decision based on the data. The key point is that big data consists of research that it is challenging to manage through traditional research tools (Mehta and Pandit, 2018).

Other terms include health data research (as used in the name of Health Data Research UK), routinely collected data (Newburn *et al.*, 2020) or data-intensive research (Aitken *et al.*, 2018; Aitken *et al.*, 2019). For Aitken and colleagues (2019), the main feature is the ability to link these data for analysis, but they should come from the same or different sources. These include:

- *"Data from patient records;*
- *Administrative data (e.g. from social care, housing or education);*
- *Data from registries;*
- *Genomic data (e.g. from biobanks);*
- *Data generated through use of apps;*
- *Social media data."*

Big data research in health can be used for multiple purposes with the aim of improving healthcare services and health inequalities (Raghupathi and Raghupathi, 2014; Zhang *et al.*, 2017). These include service management, evaluation or audit of services, statistics and exploring connections between health and non-health-related outcomes (Aitken *et al.*, 2019). Often, these purposes differ from what the data were initially collected for (e.g. healthcare or statistical purpose). In other words, big data is often used for secondary research purposes.

In this thesis, I use the term big data to describe research allowing to link routinely collected health data, which is difficult to manage using traditional analytic tools.

1.4. Conclusions

This chapter introduced this PhD thesis. It began by presenting a short background to the research and emphasising why this work was timely. It clarified the research aims. Then, it provided an overview of the research chapters. Lastly, the terms used throughout this thesis have been outlined. This will be followed now by more comprehensive literature review chapters.

As this thesis's overall aim is to explore how to involve and engage the public (especially seldom-heard communities) in big data research, the literature review consists of four chapters. First, I discuss public involvement. Second, I explore big data with a special focus on the public perspective towards it. These two separate topics require an in-depth overview before they are discussed together. Thereafter, I bring these two topics together by conducting a two-stage review. This is reported over two chapters. In one, I present the protocol for the review, and the other chapter discusses the results.

2. Public involvement

2.1. Introduction

In this chapter, I examine public involvement, one of the two key topics discussed in my thesis. First, I track its history and policy context in the United Kingdom and reflect on its position in health research. Then, I discuss in greater detail the benefits of public involvement in research and recognise a diversity of opinions towards it in the academic community. Lastly, I review what is known about how involvement activities look in practice, primarily focusing on involving seldom-heard communities.

2.2. History and policy context in the United Kingdom

This section traces the origins and development of public involvement in health services and research. Nowadays, PPI is well-established in the UK (Foot *et al.*, 2014). Yet, there have been different influences on how it came to be. These are explored through historical and policy lenses. Most of the discussions on the history of PPI start with post-1948 or even the 1970s (e.g. Hogg (2009); Beresford (2012); Stewart (2022)). This section starts at the beginning of the NHS and continues to modern times.

2.2.1. Beginning of the NHS

The creation of the National Health Service (NHS) in 1948 provided all British people access to free healthcare. The change was a positive development for patients, yet it did not establish PPI. Instead, the new system aimed to limit the influence of patients over the NHS (Grotz, Ledgard and Poland, 2020). The role of members of the public was limited, and medical professionals were put in charge. Patient organisations and charities providing services in the pre-NHS era were considered obsolete as the NHS was created, as their reason for existence was fulfilled (Grotz, Ledgard and Poland, 2020). The public influence over how the NHS is run was primarily through parliamentary elections.

Many groups did not accept this status quo and pressured the parliament to establish public involvement in healthcare services. In 1973, responding to public pressure, Wilson’s government introduced community health councils. These aimed to shift the power relationship and offer more control to the public as local residents were encouraged to meet and discuss the running of their health services. However, the public was not well aware of their role or existence (Parkinson, 2004). These evolved as PPI forums (2003), local involvement networks (2008) and local Healthwatch (2012), and thus remained as one of the PPI pillars.

2.2.2. Thatcherism

The historical and political situation kept shaping the role of public and patient involvement in healthcare services. As neoliberalism gained more political support globally, Thatcherism in Britain argued for the free market and aimed to restrain welfare, including the NHS. Interestingly, neoliberalism, through its consumerist approach, unexpectedly initiated further strengthening of patients' rights. During the Thatcherism era, some managerial changes occurred. The appointment of non-medical people as non-executive directors in NHS was allowed from 1991. This shifted the balance and gave more influence to the public.

The White Paper Working for Patients (Department of Health, 1989) advocated extending patient treatment choices. This was implemented as Major's government introduced the Patient Charter (Department of Health, 1991). As Stocking (1991, p. 1148) points out, " [chapter] *giving patients more power is not stated explicitly, but the charter should keep to achieve this*". New rights put the responsibility on the NHS to answer patient complaints promptly and receive a written response. However, the author's intention was not necessarily to empower people as citizens but rather to equate their patients' rights with consumers' rights (Plamping and Delamothe, 1991). Stocking (1991) criticises the charter for being irrelevant to seldom-heard groups such as people being homeless, people with disabilities, and ethnic minorities. In contrast to the English charter, the Welsh equivalent directly called for more patient involvement.

There was also a shift in peer-reviewed academic literature around that time. For example, the Lancet published the viewpoint of a cancer patient discussing her thoughts about joining a trial (Thornton, 1992). Afterwards, she was invited as the conference's keynote speaker to discuss patients' roles in research (Stewart, 2022).

2.2.3. New Labour

Further movement of power to the patient and public took place during the reforms of New Labour. The government's plan for health focused on improving healthcare through financial investment. In 2001, the Department of Health published a White Paper, 'Shifting the Balance of Power within the NHS', which argued for more patient involvement in shaping the services and called for a 'patient-centred NHS'. In 2005, the Department of Health (2005, p. 7) document described a patient-led NHS as a place where '*people have a far greater range of choices and information and help make choices*'. In response to these recommendations, Blair's government introduced new legislation. One of the most important laws for public involvement was the National Health Service Act 2006, requiring the NHS to include members of the public in developing and managing healthcare services. The Department of Health

(2006, p. 34) also recognised in its policy paper ‘Best Research for Best Health’ that PPI should be an essential part of all research stages, specifying them as

> ‘*priority setting, defining research outcomes, selecting research, methodology, patient recruitment, interpretation of findings, dissemination of results*.’

As New Labour was in power, the criticism towards the Patient Charter was growing. It clearly did not meet expectations of extending PPI in healthcare and proved the failure of the consumerism approach in healthcare. The study by Coulter (2011) compared survey data from six countries assessing PPI. The UK scored less than other countries. The report found that in comparison to other states, British doctors involved patients less often in their treatment decisions. Furthermore, data showed that medical professionals did not recognise patients as partners. This growing criticism urged Brown’s government to replace the chapter with the NHS Constitution for England in 2009 (Department of Health and Social Care, 2012). Despite that, the Constitution is not a legally enforceable document; it references legislation underpinning it. The Constitution states in the preamble that ‘*The NHS belongs to the people*’ and, for the first time, empowered patients to be involved in both their own treatment as well as shaping the service:

> “*You have the right to be involved in planning and making decisions about your health and care with your care provider or providers, including your end of life care, and to be given information and support to enable you to do this (…) You have the right to be involved, directly or through representatives, in the planning of healthcare services commissioned by NHS bodies, the development and consideration of proposals for changes in the way those services are provided.*” (Department of Health and Social Care, 2012)

Further reforms in the NHS were postponed as the financial crisis during Brown’s government led to austerity measures (that continued and expanded under the following governments), impacting the NHS. New Labour embedded a democratic approach in PPI. This was underpinned by the argument that if members of the public have political, social, and economic rights, then they, as users of the service, also have the right to be involved in developing and running them.

2.2.4. One-nation conservatism

Under Cameron’s government, there was a shift back to the Thatcherism era's neo-liberal policies emphasising individual rights. These undermined some of the previous reforms by returning privatisation, consumerism and shifting power from the public to medical professionals through the Social Care Act 2012 (Tritter and Koivusalo, 2013; Carter and Martin, 2016). The wording in national policy changed. The new Healthwatch was described as a ‘*consumer champion’* clearly referring to the return of consumerism in PPI (Department of Health, 2012, p. 4). The term ‘involvement’ was occasionally replaced by ‘patient choice’. Tritter and Koivusalo (2013, p. 117) argue that there is a significant difference as the involvement is “*about applying common priorities to define those options, which are different from the aggregation of individual choices”*. Local Healthwatch had mostly regional remits and thus failed to influence national policies. The survey of 150 local Healthwatch by Zoccatelli *et al.* (2020) found that only 10% of them felt that they changed national policy.

One-nation conservatism supported the growth of welfare cuts and limited the role of the public in how services should be designed (Beresford, 2020). These could have been tracked to the economic crisis of 2008 and the end of the New Labour era. However, austerity measures have developed significantly since 2010. The challenge of securing funding could lead patient-led charities to struggle financially. For example, within one and half years in 2015, over a quarter of survivor-led charities stopped working in England (Yiannoullou as cited in Beresford (2019a))

2.2.5. Where is PPI now?

Tracking the history of PPI from the start of the NHS to modern times has shown that PPI development was not linear and it could be argued that its current place in health services and research is due to multiple influences. However, the NHS approach to PPI keeps evolving (NHS England, 2017). Legislation, especially under New Labour, defined the public's right to be involved in designing, developing, managing, and evaluating healthcare services and research. The evolution of PPI shows the continuous struggle for power. The public’s right to be involved in their health services has been successfully recognised in legislation. Yet, as one-nation conservatism showed, there can be backward steps. Currently, public involvement is at one of the highest points of its influence compared to any other historical period. However, regular changes to the system, such as creating new bodies (e.g. Healthwatch), could slow down PPI expansion in the NHS as these new structures search for their place within the governance (Carter and Martin, 2016). PPI advancement could not have been achieved without generations of charities and public involvement activists, which repeatedly argued for public empowerment in healthcare services and research.

In this section, I outlined the history of PPI in the UK. I showed how it has evolved since the 1940s, with significant developments in the 1970s and the impact of successive governments. Different arguments and ideological stances shaped PPI development throughout its history. Therefore, the next section discusses these arguments for public involvement in research.

2.3. The case for public involvement in health research

The previous section identified the place of PPI within the current UK context. It can be argued that PPI is well-established within healthcare services for some, but others would still see it as a novelty (Beresford, 2020). PPI in the UK seems supported at first sight, but the reality is much more complex (Beresford, 2019b). Jeremy Taylor, Chief Executive of National Voices, representing the coalition of over 160 health and social care organisations in England, still argues that PPI is '*more of an aspiration than a programme of action; that the reality lags behind the rhetoric*' (Foot *et al.*, 2014). He names three challenges. First, PPI originates from various philosophies and historical perspectives and lacks coherent terminology. It remains challenging for professionals to know when and how to involve the public. Second, PPI challenges medical professionals' current ways of thinking, forcing them to shift their work culture profoundly. Lastly, PPI is a policy aim, yet it is not a priority. Despite these challenges, there are compelling arguments to pursue PPI in healthcare services and research. I discuss these arguments for PPI, focusing on 1) creating person-centred healthcare and research, 2) moral and legal reasons, and 3) ideological arguments. Then, I reflect on the current opinion among researchers on PPI in their work.

2.3.1. Person-centred healthcare and research

Person-centred practice is expected in modern healthcare (McCormack and McCance, 2016). Public involvement can benefit the person-centred practice as involving patients improves healthcare services and research (Lalani *et al.*, 2019; Ocloo and Matthews, 2016; Crawford, 2002; Mockford *et al.*, 2012) and brings together patients' and researchers' or healthcare professionals' priorities (Sofolahan-Oladeinde *et al.*, 2017). Patients often have in-depth and practical experiences in using healthcare services, so their lived experience should be used to advise how it could operate. Listening to patients' feedback is invaluable and if ignored, it could lead to harm. The infamous case of the Stafford Hospital scandal illustrates the danger of not involving members of the public as patients experienced poor care and high mortality rates in the Stafford Hospital. The Public Inquiry led by Robert Francis unpicked the roots of inadequate care and one of them was not listening to patients, as summarised in the report:

> *"Trust management had no culture of listening to patients. There were inadequate processes for dealing with complaints and serious untoward incidents (SUIs). Staff and patient surveys continually gave signs of dissatisfaction with the way the Trust was run, and yet no effective action was taken and the Board lacked an awareness of the reality of the care being provided to patients. The failure to respond to these warning signs indicating poor care could be due to inattention, but is more likely due to the lack of importance accorded to these sources of information." (Francis, 2013, p. 44)*

PPI could have a role in ensuring that healthcare is genuinely person-centred. For example, public contributors could advise on services and how to tailor them to patients' needs so that the public is more aware of how services operate (Mockford *et al.*, 2012). Second, patients can provide insight into what future research should explore to benefit mostly those who use healthcare services, potentially expanding on what the literature offers (Glasby and Beresford, 2006). That involvement has been embedded throughout some research organisations; for example, Cochrane Collaboration recognises the importance of PPI and runs its own Consumer Network where patients, carers and family members can provide their input on Cochrane reviews (Cochrane, 2020). However, members of the public are not equally involved at all stages of research, as the systematic review (Domecq *et al.*, 2014) showed they are more often involved in the design than in conducting research.

One of the counter-arguments against the role of PPI in being a core component of patient-centred healthcare is around conducting ethical research. An important part of ethical research is its objectivity and being unbiased. Only then it can provide an independent evidence base for practitioners to guide their work. Public contributors might have their motives and be influenced by their self-interests on how the services they use should be run. When supporting data analysis, patients might focus more on their personal experience than on data (Garfield *et al.*, 2016). Their suggestions might not benefit average patients, but a smaller group. Separating their views from participants' data is important and research shows that it is possible to achieve. For example, Hemming *et al.* (2021) utilised reflexive notes to ensure that the public contributor’s personal experiences guide the analysis but do not undermine data.

Practitioners follow the best evidence in their work. Evidence-based practice recognises the hierarchy of evidence, placing systematic reviews, and randomised controlled trials on the top of the scale, and the expert commentary, anecdotal evidence and patients’ opinions as to the least relevant. Glasby and Beresford (2006) challenged this hierarchy and claimed that patients' experiences could be as valid as any ‘objective’ research.

Ignoring members’ of the public voices because of possible bias might sound easy, but one could argue that this would be immoral and undemocratic. Both healthcare services and research should seek to make a difference in people's lives. Those benefiting could have a right to participate in how (and if) it would benefit them. This argument assumes that if the NHS is publicly funded, taxpayers should exercise control over it both indirectly (through elected representatives) and directly (through PPI). The public ownership argument has been recognised among healthcare research funders. Many require PPI at the application stage (e.g. NIHR), thus recognising PPI as one of the core elements of good research. Involving the public from an early stage is supported by the literature as PPI can be beneficial from the moment of study design (Staniszewska *et al.*, 2007). The UK Policy Framework for Health and Social Care Research (NHS, 2020) reaffirms this as one of its principles:

> "*Patients, service users and the public are involved in the design, management, conduct and dissemination of research, unless otherwise justified.*"

Through a person-centred approach, PPI could bring financial benefits to NHS. As the population ages, healthcare becomes more expensive to cover the needs of older people. Foot *et al.* (2014) claim that successful PPI can make services more sustainable and cost-effective. First, if patients are involved in managing their healthcare, it could reduce their reliance on public services. Second, the NHS should be able to provide appropriate service when medical professionals are aware of their patients' needs and thus provide them with what they need and when they need it. Thus, this suggests that it could reduce pressure on the NHS in the long term.

2.3.2. Moral and legal reasons for PPI

Ethics in healthcare services and research could be summarised as "first, do no harm". In the research context, participants have the right to receive information about the positive and negative consequences of participating in the study to make an informed decision if they want to participate. Involving public contributors who are also community members could assist researchers in developing recruitment materials. They can also help understand future participants' backgrounds and needs, such as culture, language, beliefs and life circumstances, to ensure researchers consider them during the research. Conducting a review might give the researcher an insight into what a particular community experiences but cannot replace a conversation with them. There are two reasons behind this. First, previous research might not be generalisable to that specific group. Second, the studied community might disagree with the results of previous studies.

Regarding public contributors' involvement in individual projects, members of the public could sit on ethics or governance committees to bring a lay perspective when reviewing research proposals. Research Ethics Committee - Standard Operating Procedures (NHS, 2019) requires at least one lay member to be present for meetings to be quorate. However, there remains a difference in which type of projects public contributors can get involved. Boylan *et al.* (2019) show that in academia a perspective that remains is that people can get involved more easily in qualitative research rather than quantitative research. This is because qualitative research could be associated with their lived experiences and quantitative might require specialist knowledge. Among those who support public involvement in research, not all researchers consider all projects appropriate for PPI at all research stages (Garfield *et al.*, 2016).

Involvement standards are also included within legislation and put binding requirements on NHS to include the public, for example, the already mentioned Health and Social Care Act 2012. Medical professional codes as well include references to PPI in individual care. For example, the professional standards of practice and behaviour for nurses, midwives and nursing associates (NMC, 2018, pp. 6-7) require them to listen to patients and

> "*work in partnership with people to make sure [they] deliver care effectively* (…) *encourage and empower people to share in decisions about their treatment and care.*"

Working in partnership with the public could mean that the public should be able to understand research findings without specialist knowledge. Journals such as the BMC Research Involvement and Engagement require a lay summary of each paper. Patient Included journal charter also encourages journals to include patients on the editorial boards, as paper co-authors, and peer-reviewers, and publish all papers as open access (Patients Included, 2020). The paywall is an obstacle to members of the public who do not have institutional access. If, ethically, the public has the right to read research findings, these need to be available free of charge.

This subsection has shown that there exist moral and legal reasons supporting PPI in research. However, these are underpinned by ideological arguments and I discuss them in the following subsection.

2.3.3. Ideological arguments for PPI

Two main opposing ideological origins of PPI may be broadly identified, and these are politically inherited. First, the government, researchers and funders want to build trust between health services and the public. On the other hand, service users and patient organisations aim to explore how much health services can be placed under public control (Beresford, 2003; Beresford, 2002). These two approaches are identified as distinctive ideologies of involvement: managerial & consumerist and democratic.

The Thatcherism era of the private market and individualism influenced how NHS should operate. Underpinned by these ideas, managerial and consumerist approaches place patients as consumers (Beresford, 2003). Thus, it suggests that NHS is a business (Harrison, Dowswell and Milewa, 2002), so dissatisfied customers (patients) have the right to raise their complaints concerning their care. The approach derives from the state level, and Beresford (2012) claims this is the UK's dominant model as it continued through the New Labour and

One-nation Conservatism eras. The main focus of consumerism is a consultation with the public on the policy and planning, but how this feedback is utilised remains in the power of healthcare professionals and management (Beresford, 2002).

In contrast, the democratic approach links the individual rights of patients as citizens. The NHS is a publicly owned institution, so the ownership and decision-making should extend to members of the public. The public should participate and decide on changes. Proponents of this approach argue for redistribution of power and thus what Hogg (2009, p. 59) calls a '*tackling of a democratic deficit'* of the NHS. The public should be more involved in how services are shaped and what research is conducted. That involvement would give patients more say and influence on their lives and the service they experience. The sources of it can be traced to the origins of representative democracy, where the public is represented in the state's governance (Harrison, Dowswell and Milewa, 2002).

Supporters of both approaches agree on PPI's need but not what it means in practice (Beresford, 2002). Approaches co-exist nowadays, causing clashes regarding power redistribution within the NHS. Hui and Stickley (2007) suggest different approaches can be identified through how people use language and rhetoric. Therefore, based on their discourse analysis of literature and health policies in the UK, they recommend raising awareness of how subtle language changes can influence public contributors and healthcare professionals. However, the rhetoric can also be related to professionals and researchers' view towards public involvement, for example, if they are supportive towards public involvement in their work. The diversity of these views is explored in the next subsection.

2.3.4. Researchers' views on PPI

Previously, I deliberated on arguments supporting PPI in research. However, the strong case for PPI does not mean that all researchers are equally supportive or always involve the public in their research. Becker, Sempik and Bryman (2010) surveyed 251 social policy researchers and found that three-quarters of them consider public involvement very or fairly important during all research stages, with the remaining participants finding it of little relevance. Giving power to the public means that researchers' influence exists alongside that owned by members of the public. Some professionals believe that the public has a limited understanding of what they do (Becker, Sempik and Bryman, 2010), simply do not care, patient organisations are insignificant and thus should not be involved in any decisions on health services and research (Mouton Dorey, Baumann and Biller-Andorno, 2018). However, researchers could rectify this 'lack of knowledge' among the public by offering training, leading to a better understanding of academic jargon and higher interest in research among members of the public (Simpson *et al.*, 2006). Workshops provide members of the public with skills to contribute in all research stages, e.g. data analysis (Garfield *et al.*, 2016). Hemming *et al.* (2021) involved public contributors with lived experience in the thematic analysis of interviews. This was achieved successfully through three hours of training designed by the doctoral student (the author of the paper), which included asking the public contributor to code sample transcripts as an exercise. The success of that project was possible thanks to building a strong relationship between the public contributor and the researchers as he was involved from the beginning and thus, this ensured that the power imbalance between the researcher and public contributor was limited. On reflection, Hemming *et al.* (2021, p. 6) claimed that training of the public contributor in the thematic analysis did not "*dilute*" his perspective as a person with lived experience. Instead, it seems that including the public contributor in the study offered a more nuanced analysis of interviews.

Becker, Sempik and Bryman (2010) suggest PPI 'support' typologies among researchers as advocates, cautious advocates, agnostics and adversaries. Advocates view PPI positively and offer solid, if not unconditional, backing. They recognise the benefits of PPI (e.g. improved quality, better dissemination, participant recruitment, study design and ensuring its relevance). Cautious advocates generally support PPI but believe it depends on the project and the public's ability to contribute meaningfully. Agnostics would disagree with PPI in practice. They believe that PPI is ideological and does not add any additional value to their work. Adversaries would have hostile views towards PPI.

PPI's organisational attitude can support or discourage these perspectives, primarily if senior management supports PPI culture (Boylan *et al.*, 2019). One of the interviewed researchers in Boylan *et al.* (2019, p. 726) study said

> "*Quite honestly, the very senior people think this is a waste of time and a box-ticking exercise. And a lot of what I've been doing I've been keeping it quiet, because I don't want anyone to tell me that I'm wasting my time*."

Researchers who involved the public in their studies reported how beneficial it was to them. Garfield *et al.* (2016, p. 7) interviewed researchers who involved members of the public in qualitative analysis. One of the study participants said that it "*added a different dimension to some at least of the findings.*"

In this section, I outlined the arguments supporting public involvement in research. I especially focused on person-centred care, moral, legal and ideological arguments. These showed a strong case for public involvement, but this support might be based on different ideologies. Thus, the overarching goal of involving the public might be shared, but what means are needed to achieve it or how it could look might differ. I also recognised that there is some opposition to public involvement. I discussed this division when exploring researchers' views and experiences when involving members of the public in their work. In summary, the section looked into ‘why’ researchers should involve the public and showed that there are different perspectives among academics. Therefore, the next section will focus on ‘how’ and discuss the involvement practices and what makes people interested in public involvement.

2.4. Involvement in practice

PPI models on how to involve members of the public give recommendations for reaching communities (including seldom-heard) in health services planning (Snow, Tweedie and Pederson, 2018), children and young people in research and health services (Brady, 2020) and the general public (Grotz, Ledgard and Poland, 2020). What all these models have in common is that they argue against perceiving PPI as a linear activity. Although PPI models could guide researchers, they tend to oversimplify the process (Beresford, 2019b; Beresford, 2021). In this section, I analyse what is known about why people get involved in PPI and how to support their motivations based on these models and public involvement literature. Then, I reflect on why involving seldom-heard communities is important. Finally, I explore what is known already about conducting involvement with these communities.

2.4.1. What makes people enthusiastic, interested and committed to PPI?

Patients and the public who contribute to PPI are volunteers (Cochrane, 2017). Thus, the literature on PPI and volunteering could offer insights into why people get involved. Their participation is influenced by personal and social motivations (Turk *et al.*, 2020; Tarpey, 2006). Although the following list is not comprehensive, it provides an overview of why the public might want to get involved as public contributors. Understanding why people get involved is limited; the literature review from 2006 found that most PPI studies explored why people avoid involvement activities, and papers did not explore motivators in detail (Tarpey, 2006). However, despite this review being a couple of years old in this evolving field, more recent evidence adds that PPI requires additional time and resources (Boden *et al.*, 2021).

Personal motivations primarily focus on what individuals gain from participation. Public contributors can do something new, learn new skills (e.g. around research), expand their knowledge about their sickness, build confidence (especially if their illness was prolonged) and earn money through financial reimbursement (but this is not always the main incentive) (Newington *et al.*, 2022). They might feel that involvement would allow them to voice and decide about the service - especially if they or their family members were frustrated with the care they received. Individual reasons for participation would also include ‘wrong’ motivations, such as shaping service to make it more comfortable for themselves rather than average users.

Social motivators often refer to more altruistic reasons for contributing something back to the community. Public use their acquired knowledge of health service to build more patient-centred service- especially if they had poor prior service experience (Staniszewska *et al.*, 2007). Patients want to offer something back to others to show gratitude for the support they received. They might be already active in local charities representing patients of one specific illness (Tarpey, 2006). Researchers often recruit them directly when approaching local charities for advice. People might feel this is their responsibility to represent their community.

Participation in PPI can sometimes be overwhelming with the amount of work required (Staniszewska et al., 2007). This issue emerged in Turk et al. (2020, p. 8) study of interviews with public contributors and researchers. Researchers were concerned that asking public contributors to sit in the room with unknown experts and do too much would be burdensome - thus potentially discouraging them from being involved. One of the public contributors said that researchers should not be too concerned about these issues as there was a "*huge well of enthusiastic altruism.*" This shows that researchers should keep involving members of the public in their research but be honest about the time commitment from the start. However, how to achieve that remains unknown.

2.4.2. The importance of involving seldom-heard communities

Grotz, Ledgard and Poland (2020) recognise that people's views and experiences might differ from what researchers believe. This might be especially the case when involving seldom-heard communities. Including all communities in public involvement remains challenging (Flynn, Walton and Scott, 2019; Ocloo and Matthews, 2016; IPFCC, 2018). Still, it is one of the fundaments for successful involvement (Harrison *et al.*, 2019b; INVOLVE, 2012a; Aitken *et al.*, 2019). The lack of including seldom-heard communities can increase health inequalities if the views of these groups are ignored (Hogg, 2009). However, this should not be a tick-box exercise (INVOLVE, 2012a). Researchers and policymakers must include diverse voices to ensure services can effectively respond to all communities' needs (Phillimore, 2011).

Recording participants' demographics is crucial to ensure and evaluate if seldom-heard groups are included. In many PPI projects, this is still not the case. The Health Research Authority (HRA, 2021) explored the public involvement matching service, which linked researchers putting Covid-19 related research projects to reach patient groups for feedback. Their report found that most groups did not collect demographic data about public contributors, such as ethnicity (one-third missed it). This leaves the question if findings from these public involvement activities would benefit seldom-heard groups. Ethnic minorities such as Polish have a history of lower trust in vaccinations than the British population. In the study taking place before the Covid-19 pandemic, Polish participants often compared their experiences of vaccinations to their home country. Therefore, they expected a doctor to conduct pre-vaccine checks and not a health visitor. (Gorman *et al.*, 2019).

Including seldom-heard communities does not equal ensuring representativeness. The PPI participants can be diverse but not representative. Parkinson (2004, p. 373) summarised that point by saying, "*One can only have descriptively representative samples, within certain confidence limits, not descriptively representative individuals*".

In this section, I recognised the importance of involving seldom-heard communities, but the next subsection will outline what is already known about involving these communities.

2.5. Involving seldom-heard communities

A body of literature focuses on the recruitment of seldom-heard voices as study participants (Bonevski *et al.*, 2014). However, the literature on their inclusion in public involvement remains limited (Harrison *et al.*, 2019a). Researchers are still looking for new ways to understand how to effectively involve seldom-heard groups (Volkmer and Broomfield, 2022). Seldom-heard voices could be motivated to participate in public involvement, but researchers need to find innovative ways of reaching them (Tarpey, 2006). This section discusses what is already known in the literature on this. The PPI literature has many anecdotal and descriptive examples of engaging with the public and patients, including those from seldom-heard communities. However, Steel (2005b) argues that covering all aspects of public involvement these communities might face is infeasible. There is no single path to achieve successful involvement (Grotz, Ledgard and Poland, 2020; Brady, 2020), so I discuss overarching principles of good practice instead. These focus on two broad themes: inclusive environment (ensuring welcoming involvement culture) and capacity building (trust, building relationships and developing skills of public contributors).

2.5.1. Inclusive environments

Providing an inclusive environment could ensure that public contributors have the confidence and opportunity to contribute fully to any involvement activities. Researchers need to use lay vocabulary as jargon and specialist language could marginalise anyone unfamiliar with that vocabulary (Steel, 2005b). However, limited evidence shows how to use lay language in complex health research topics (such as big data research). Additional challenges come for those who are not native speakers (especially if their English knowledge is limited), as they might require translations and interpreters. It would be ethical for a researcher to obtain informed consent (Morrow, Boaz and Sally, 2011). The lack of translated materials remains an obstacle in healthcare services. In a study among Polish women on vaccinating their children in Scotland, one of the participants pointed out that '*it's hard for me… because you* *have to beg for information…*'. (Gorman *et al.*, 2019, p. 2744). This shows that accessing information in the first place can be a challenge for those whose English is the second language. Although the authors of that paper point out that these materials were available in Polish, participants did not find them, so they relied on other sources of information such as Polish relatives and friends, media (including social) and the internet. This questions if translated materials could be used for involving non-native English speakers.

Facilitators of the public involvement activity should ensure that participants feel comfortable contributing to the discussion. Grotz, Ledgard and Poland (2020) broadly identify three principles to achieve that aim. First is mutual respect, where public contributors are treated equally and with respect (even if there is disagreement) by other group members and facilitators. How that would look in practice and could be achieved remains underexplored.

Second, accessibility in terms of access (e.g. for people with disability) or communication (e.g. access to interpreters). Although interpreting can make some activities difficult to conduct (Gilbertson *et al.*, 2021) researchers should be prepared to adopt involvement methods based on each participant's needs and confidence. Broomfield (2022) involved public contributors who use augmentative and alternative communications and showed that everyone could contribute by adjusting the way involvement activities were set up (e.g., using photographs or illustrations). Cotterell and Paine (2012) gave an example of a public contributor with an ongoing health problem who needed access to oxygen and had to take regular medication not to experience pain and discomfort. These two examples also show adjustments can sometimes require further resources regarding the researcher's time and finances.

Third, a public involvement activity should take place in a safe space where public contributors are confident to express their views without feeling that it might impact the type of service they receive in the future. The power relationship between public and professional members during involvement activities can affect the discussion's quality (Snow, Tweedie and Pederson, 2018). Anonymity and independent facilitation could alleviate these feelings. Also, researchers should ensure that all research terms and concepts are explained to people in accessible language (Romsland, Milosavljevic and Andreassen, 2019). However, there is no clear guidance how to ensure that language is accessible in complex health topics.

Public involvement activity can be a conversation, similar to a focus group. However, creative approaches such as artwork can make PPI more approachable and accessible to seldom-heard groups (Tierney *et al.*, 2021; Gilbertson *et al.*, 2021). Producing artwork offers a unique way to discuss public contributors' views on research and could offer "*an eye-catching and accessible way to disseminate these potential ideas for future research*" (Tierney *et al.*, 2021,

p. 4). Eccles *et al.* (2018) suggest ‘mobile workshops’ as an alternative to members of the public, who they describe as ‘the un-usual suspects’ (people who might not be from a seldom heard community but are not getting involved because of time commitment). These workshops were organised in places where people were already (e.g. work or social group) and took only 30 minutes. Further ideas come from science engagement activities such as PPI Café or PPI Hawkers (Luna Puerta and Smith, 2020). Similarly, to science café, researchers discuss here PPI issues with the public in an informal manner, often accompanied by both alcohol and non-alcoholic beverage. Whereas PPI Hawkers was tested in Singapore, researchers joined people at their tables during meals at traditional food outlets called Hawkers. Broomfield *et al.* (2021) showed that people with communication difficulties could benefit from creative approaches, such as audio-visual media (recording agenda as audio and uploading online, using scales or Lego as a way for participants to produce feedback). They also had six other meetings and different activities offered in a clockwise manner (by the last meeting they understood the project and built confidence around it) (one activity per meeting), this project showed the potential of helping people with communication difficulties to understand what the research project is about and provide their feedback to it.

A researcher should not presume what participants' needs might be based solely on literature or their assumptions but discuss it (where possible) directly and implement any suggestions (Morrow, Boaz and Sally, 2011; Starbuck *et al.*, 2020; Nguyen, Cheah and Chambers, 2019). Even if considered by research as transferable, the experiences of other groups might not reflect the needs of different communities (Dixon *et al.*, 2018). These discussions also provide an opportunity to build capacity for successful future involvement, even if these public involvement activities are short (Eccles *et al.*, 2018).

2.5.2. Capacity building

Research shows that healthcare professionals found that involvement is most successful when it happens both formally and informally (Snow, Tweedie and Pederson, 2018) as it allows understanding of the beliefs and culture of the people researcher wants to involve (INVOLVE, 2012c).

PPI requires planning and time (Romsland, Milosavljevic and Andreassen, 2019). Ongoing relationships with patient groups can be beneficial. However, little is known about how to establish the relationships for PPI. These relationships could assist researchers when there is limited time to design research (e.g. as it happened at the beginning of the Covid-19 pandemic). When the COVID-19 pandemic started, the HRA (2021) found that only 22% of research applications in March-April 2020 had a PPI component. This was in contrast with 80% of pre-COVID-19 applications. However, the new matching service developed by the

HRA allowed researchers to engage quickly with the right public contributors; thus, by August 2020, 85% of applications included public contributors.

More research organisations have established a pool of volunteers interested in participating in public involvement activities (Grotz, Ledgard and Poland, 2020). For example, NIHR offers members of the public opportunities to join local research groups as public contributors (NIHR, 2021b).

Researchers should understand the context and views of the people they involve. Snow (2018) suggests that when working with people from diverse gender identities, participants could be asked what pronoun they wish to use at the start of the meeting. The second example was when working with Aboriginal communities. The researcher had to adapt to the matriarchal nature of the community they were involved with, so they understood the importance of elders and other cultural traditions. INVOLVE (2012c) report adds that each group might have specific needs, and those can be related to the time of year; for instance, the involvement of faith communities should be avoided on their holy days. Despite these suggestions, not all seldom-heard communities have been explored; for example, there is no research exploring how to involve the Polish community.

Grotz et al. (2020) recognise two approaches to reach potential participants from specific groups. They called them ‘outreach’ and ‘in reach’. The former encourages going to the communities to establish relationships through activities such as stalls or presentations. Researchers aim to develop their links with the community in the latter approach by offering them new access to resources and additional activities.

Participation in involvement activities could include direct and opportunity costs (Snow, Tweedie and Pederson, 2018). Direct expenses cover travel costs (time and money) or childcare. The latter refers to missed social or work opportunities as participants dedicate their spare time to attending the session. Snow et al. (2018) point out that lower socioeconomic status is linked with more substantial barriers to involvement as these groups would have limited resources. The middle class can afford to be involved and thus might have a higher chance of involvement. Reimbursement for time is an essential issue for potential public contributors. The research project should cover any direct cost of involvement, and these should be paid in advance as not everyone can wait to be reimbursed (Starbuck *et al.*, 2020). Other ways of making participation accessible to seldom-heard voices could include flexibility about locations and times; for example, Simpson *et al.* (2006) argue that shorter sessions could increase involvement from those who are carers. Public contributors gain opportunities to learn new skills, build new relationships and improve their CVs. However, these personal benefits would be less of an incentive for those struggling financially (Snow, Tweedie and Pederson, 2018).

Challenges around involving seldom-heard groups are often similar to the involvement of the general population, but they might be more visible (e.g. direct cost) than the majority of the population. Thus, research findings from the involvement of seldom-heard groups could be transferable to the general population.

In this section, I discussed what is already known about public involvement in health research. Then, I emphasised why involving seldom-heard communities is important for researchers. Thereafter, I focused the discussion on what the public involvement literature says about working with seldom-heard communities. There is growing evidence of how to involve these communities. However, none of the studies mentioned in this section specifically refer to big data research that is of interest to my thesis. This questions whether big data researchers would share similar experiences or face specific involvement barriers.

2.6. Conclusions

This chapter explored public involvement in healthcare services and research. It presented the PPI history and arguments for involving the public. I showed that there are different sources of support for public involvement, and there remains division among researchers if, why and how the public should be involved in research. Understanding researchers' perspectives and experiences of PPI requires further attention. Hence, one of my studies focused on researchers' experiences.[4] Then, I explored what we know so far about involving members of the public in research and emphasised the need to include seldom-heard communities. Lastly, I looked at existing evidence about involving these communities and recognised that not all seldom-heard communities have been explored (e.g. Polish community in the UK). Therefore, one of my studies specifically focused on ethnic minorities.[5] The literature on public involvement is growing and provides some suggestions for researchers, but it seems not to include anything about big data research. This gap requires further investigation, and I explore it in a later chapter when looking at public involvement in big data research.[6] However, before I look at these two issues together, the next chapter focuses on exploring literature around big data research.

---

[4] See chapter 10 at page 174.

[5] See discussion at the choice of participants at page 105.

[6] See chapter 5 for this discussion at page 72.

3. Big data

3.1. Introduction

In this chapter, I focus on big data research. The chapter provides the context where PPI in big data research occurs. First, I discuss the public views towards big data research, including their expectations. Second, I provide an overview of public concerns when researchers reuse patient data. Then, I present three UK cases of big data initiatives that caused public controversy. Finally, I consider what is known about social licence's role in big data research.

3.2. Public views towards big data research

Public awareness of big data research remains limited. The Wellcome report by MORI (2017) found that only 33 per cent of the survey participants had a great or fair understanding of how NHS is using their data. This further went down to 18 per cent for university researchers and 16 per cent for private companies. The systematic review by Stockdale, Cassell and Ford (2019) found that the public is mostly aware of data collection, but data-sharing practices for research remain low. Despite the majority of the public being unfamiliar with big data, there has been extensive research on their attitudes towards big data research (e.g. see Kalkman *et al.* (2019b) or Aitken *et al.* (2016)). The public mostly supports it (Aitken *et al.*, 2016) but remains concerned, for example, about data breaches and data misuse (Kalkman *et al.*, 2019b). Public opinion on big data initiatives is alike in different countries; for example, Tully *et al.* (2020) identified similarities between public views in Scotland and Sweden towards reusing their medical data, with both nations being equally open to it.

However, not all community groups might have the same view. The study by Hill *et al.* (2013) indicates that males and the older population are more open to sharing their medical data. The systematic review of genomic data sharing suggests that older populations are less worried about their data privacy (Shabani, Bezuidenhout and Borry, 2014). Another systematic review (Stockdale, Cassell and Ford, 2019) found differences in the support between different age groups, people of various education levels, socioeconomic levels and ethnicity. Interestingly, they found that older generations would be more worried about data security than younger people. These contradictory findings indicate that more research is needed into public perspectives on big data research.

Stockdale, Cassell and Ford (2019) also suggest that non-white British are more likely to be concerned about the invasion of their privacy and, thus, less supportive of data sharing than White British. This could be explained by seldom heard communities' trust levels in NHS. Bailey Wilson *et al.* (2022) found that Black and South Asian communities in the UK have lower trust levels in NHS and might have more concerns about how their data would be used or protected. This lack of trust could have been caused by historical and societal causes when these communities experienced racial discrimination. This suggests they might fear data sharing could lead to further racial discrimination rather than health improvement (Bailey Wilson *et al.*, 2022). However, ethnic minorities are not the only ones to remain sceptical if the public would benefit from big data research. The general public is concerned that there would be enough political support to deliver real change to the community (Aitken *et al.*, 2018).

The public might understand the advantages of big data but also expects that the research would offer benefits (Howe *et al.*, 2018) and thus are happy for their data to be reused for the common good (Stockdale, Cassell and Ford, 2019). What the benefit of big data research should be, remains broad among members of the public (Aitken *et al.*, 2018), but Aitken *et al.* (2016) recognise it as one of the key conditions for public support. Stockdale, Cassell and Ford (2019, p. 9) define the common good as "*consist[ing] of the collective public health benefits brought about by the improvement of the services, practices and methods of healthcare through secondary uses of data*". There seems to be a consensus that the term 'public' should be as inclusive as possible and individuals should share benefits of big data research, broader society and specific groups that research might target (Aitken *et al.*, 2018). The expectation of benefit is as important as the public experiencing that positive change (Howe *et al.*, 2018). How communities perceive this benefit could differ. Black and South Asian communities are convinced that big data research could offer novel healthcare solutions; however, they remain unconvinced that it would benefit directly their communities as they question if the NHS considers them a priority (Bailey Wilson *et al.*, 2022).

Health data could be linked with non-health data for research purposes. Here, again the public perspective differs on which other databases they would feel comfortable for their medical data to be shared with. Jones *et al.* (2022b) conducted an online survey to explore public views on sharing health data for research purposes without explicit consent. Their findings suggest that the public would feel the most comfortable linking their medical data with information held by the universities (72.8% support). It was followed by data around social security and pensions data (56.6% support) and only 17.3% for private companies.

Media reports on big data research could influence public opinion. Portland Communications (2021) conducted an analysis, for the Understanding Patient Data[7], of news stories about health data in the UK media. Their findings suggest that stories that report data breaches or negative news receive the most media attention. For example, during the Covid-19 pandemic, there was a spike in stories reporting that health data could be reused by police or insurance

---

[7] Understanding Patient Data is an organisation that aims to share information about how health data is used and raise public views around data sharing with stakeholders. It is hosted by the NHS Confederation.

companies. However, at the same time, they found that news stories remain mostly positive by sharing the potential benefit of big data research (e.g. how it could help with the Covid-19 pandemic) but do not explore the processes behind that research. Consequently, this media coverage could have a limited impact on improving public awareness of what big data research looks like.

From the public perspective, big data research could offer benefits and novel healthcare solutions. However, at the same time, they would consider their health data to be sensitive information (Taylor, 2014). This sensitivity raises some concerns about big data research among members of the public, which I discuss in the next section.

3.3. What are public concerns around big data?

The previous section showed that the public acknowledges the benefits of big data research and that the majority supports data reuse, but this is often conditional (e.g., by ensuring safeguards) (Aitken *et al.*, 2016; Kalkman *et al.*, 2019b). The public concerns may be broadly categorised around three main themes: lack of consent, confidentiality, and data misuse. I discuss these concerns and then reflect on what could be done to overcome them.

3.3.1. Lack of consent

One type of consent for big data research is the opt-out. This means that patients' data is automatically included unless they opt out of the database. The opt-out model is the opposite of the opt-in model, where people need to provide direct consent for researchers to reuse their data. When medical data is collected, it might not be clear how it might be reused and if at all. Therefore, obtaining consent when patient data is acquired (e.g. during a medical appointment) is impossible. Having the opt-out system ensures that more data is included and avoid bias. Bias in big data research could mean that only some sections of the community are included in the database. The opt-out model could cause public controversy among some members of the public. The systematic review by Garrison *et al.* (2016) found that the public’s preferences remain mixed in the literature on biobanking, broad consent and data sharing, as in some studies, participants preferred the opt-in model. In contrast, in others, the opt-out model was more appropriate or sometimes could be a preferred option. That divide is well illustrated in the qualitative study by Hill *et al.* (2013), where all participants agreed that they would consent for their medical data to be reused for research if asked. However, their opinions around reusing their medical records without prior consent were evenly split into those who were open to it and those who believed they should be asked first. One option to offer a more acceptable model of consent for those members of the public who would like some form of control could be the Dynamic Consent (Spencer *et al.*, 2016). It offers a more accessible way for members of the public to opt out through an electronic system that could be updated anytime. At the same time, the Dynamic Consent has the potential for big data research to include as broad participation as possible. However, there remains the risk that those with poor access to technology or lacking IT skills would be incapable of using the application (Spencer *et al.*, 2016).

3.3.2. Confidentiality

The privacy and confidentiality of patient medical data are important to members of the public (Aitken *et al.*, 2016; Howe *et al.*, 2018). Some data, such as ethnicity or if a patient is HIV positive, would be more sensitive than others (Howe *et al.*, 2018). The ability to reidentify individuals is a particular concern (Kalkman *et al.*, 2019b). However, even if data is deidentified, privacy remains a concern for the public (Shah *et al.*, 2019). Big data often use pseudonymisation, where individual identities are removed. In theory, the individuals could be reidentified if someone has access to the pseudonymisation code. The public perspective on data security (e.g. the existence of privacy measures) could increase members of the public willingness to share their data (Kalkman *et al.*, 2019b). The confidentiality issue relates to the potential data misuse that I discuss in the next subsection.

3.3.3. Data misuse

The public is concerned that there is a possibility that their personal data might be misused. This includes an unauthorised publication of personal information or the reuse of it to potentially harm a member of the public (e.g. by reporting their misconduct to authorities like social services) (Manhas *et al.*, 2015). However, the main concern is that private companies could misuse data (Kalkman *et al.*, 2019b), even if it is an unintentional breach (Bailey Wilson *et al.*, 2022). The public would be less supportive of a private company's big data projects if there is no explicit benefit to the general public (MORI, 2017). There are questions about how altruistic private companies are when they use patient data (Manhas *et al.*, 2015). This especially relates to concerns about how insurance companies could use data against members of the public (e.g. influencing their covers) and marketing companies to target ads at specific individuals (Hill *et al.*, 2013; Manhas *et al.*, 2015). Only 26 per cent of survey participants in the Wellcome report supported sharing anonymised health data with insurance companies (MORI, 2017). The public concerns could be considered justified. Access to data for some companies would be a gold mine (Abbasi, 2022). The British Medical Journal investigation found that private companies, NHS commissioners and universities numerous times violated data-sharing agreements (Oxford, 2022). The NHS Digital at the time of investigation[8] was responsible for checking if the safeguarding policies were followed and

---

[8] NHS Digital merged with NHS England on 1st February 2023.

could suspend the provision of data. However, they decided not to use any enforcement actions but rather work with organisations to improve their activities. There were concerns raised by former lead for the Understanding Patient Data Natalie Banner that "*these breaches will damage public trust that data are being handled safely and securely*."

In contrast to public companies, there is an indication that the public support for NHS and the medical profession to reuse data is higher. Spencer *et al.* (2016), in their qualitative study with patients from the local rheumatology outpatient clinic and PPI health research network, found that participants had a high level of trust in healthcare around storing patients' data confidentially. This suggests that close relationship with healthcare professionals improves the public trust in their handling of that data.

The public wants data protection measures to be implemented (Howe *et al.*, 2018). They could be more supportive of big data when safeguards such as anonymisation and data encryption are established (Hill *et al.*, 2013). However, no single safeguard could reassure the public (MORI, 2017), and to some, no protection system is fully secured (Spencer *et al.*, 2016). Offering a way to control patients' data to the public could empower them (Howe *et al.*, 2018). More control (e.g. safeguarding) could improve the public support for data reuse by private companies (MORI, 2017). Shah *et al.* (2019) identified the relationship between happiness to share data with private companies and the need for control. The public has confidence that researchers hold accountable those who breach public trust and misuse data for personal gain (Kalkman *et al.*, 2019b). However, Dixon-Woods and Ashcroft (2008, p. 389) argue that creating any regulations around big data research is "*Janus-faced*". It could reassure the public that big data research offers benefits and is conducted in an acceptable way. On the other hand, the need to create regulations creates unease among members of the public. Some might be concerned about why these are put in the first place.

Even if all safeguards are implemented, some members of the public would still prefer the opt-out model rather than the opt-in one (Hill *et al.*, 2013). In this section, I discussed public concerns around big data research. The following section illustrates how these concerns played out in previous big data initiatives in the UK.

3.4. Public controversies around big data research

New big data projects in the UK caused public controversies and could have negatively influenced how the public might react to these kinds of initiatives. I now discuss three high-profile cases from England to show the consequences of poor rapport between researchers and the public. These include care.data, Google Deep Minds and the General Practice Data for Planning and Research.

3.4.1. Care.data

The project was announced in 2013 by the NHS Digital[9] and aimed to be England's first single database of all medical data. If successful, it would have stored all the medical history of NHS patients in one place. Thus, it would allow researchers to easily access that information for big data research. However, multiple controversies appeared around the project. These can be divided into transparency and communication issues. Care.data was unclear (and thus lacked transparency) on who should have access to the data and if members of the public could know who actually has their data or what it is being used for (Ramesh, 2014). Taylor (2014) went a step further and questioned if NHS would even have had to respect patient disagreement for their data to be shared as there was no clear legal requirement for it. The further public concern was that private companies (including insurance) would be able to acquire these data for their work. As data would be pseudonymised, in theory, some companies could match them with their own databases and potentially identify individual patients (Ramesh, 2014). In terms of communication, the care.data was poorly promoted, so the public's awareness of the initiative was limited (Taylor, 2014). The promotion campaign aimed to improve that situation was led by the NHS and delivered a leaflet to over 25 million homes in England titled "*Better information means better care*." The leaflet was criticised for not being clear enough as it did not mention care.data by name and encouraged the public to reach their GP practice for more information (Hays and Daker-White, 2015). The information provided was biased as it mentioned only benefits, but no risks and the opt-out form was not included (Carter, Laurie and Dixon-Woods, 2015). Afterwards, a poll by ICM Research showed that only 29% of surveyed adults recollected getting the leaflet and around 45% were unaware of the plan in 2014 (Vallance, 2014). Because of poor communication and unclear transparency, anti-scheme campaigners called it a "*disaster just waiting to happen*" (Ramesh, 2014). Both proponents and opponents of the project shared some of these concerns (Hays and Daker-White, 2015). The programme was postponed and finally cancelled in 2016.

3.4.2. Google Deep Minds

Deep Mind is owned by a private company, Google. On one of their health-related projects, the company worked with the Royal Free London NHS Foundation Trust to create a mobile application around kidney disease. For the application to work, Google Mind had access to linked patient records. However, there was no kidney-specific database in the Trust, so Google had access to all patients' data. The application received positive feedback from healthcare professionals, and nurses felt it could save up to two hours of their daily work (Royal Free London NHS Foundation Trust, 2017). The public controversy was around the amount of

---

[9] NHS Digital was known that time as the Health and Social Care Information Centre.

access Google received. The trust shared the medical data of 1.6 million patients that included information about their admissions, pathology, critical care and sensitive information such as if a patient is HIV-positive. The data were not pseudonymised or encrypted. There were also concerns about why there was no ethical approval around the project and if patients were made aware of how their medical data could be used (Hodson, 2016a; Iacobucci, 2017). Public engagement and transparency were limited. The company did not share the depth of the partnership and their access to medical data till the media inquiry (Hodson, 2016b). After the investigation, the Information Commissioner's Office in 2017 found that personal data transfer to DeepMind breached data protection legislation. However, the public body did not issue any fines. As a part of the response from Google, the company promised more proactive engagement with the community and worked with public contributors to create an involvement strategy (Suleyman and King, 2017). In 2021, Google announced the closing of the application project (Lomas, 2021). Currently, there is the potential of a class action-style lawsuit in the UK against Google over Deep Mind's data use (Simister, 2022).

3.4.3. General Practice Data for Planning and Research

In 2021, NHS Digital announced plans for a new big data initiative called the General Practice Data for Planning and Research (GPDPR). The idea was to link (pseudonymised) patient medical data of 55 million patients in England and make it available to researchers. From the start, the project experienced public controversy and was even called by critics the "*care.data part 2*" (Crouch, 2021). Similarly to the care.data project concerns were raised around transparency (especially around informed consent) and confidentiality of patient data (O'Dowd, 2021). What was new in this initiative was that the time for the public to decide to opt out of this new database was restricted (Shaw and Erren, 2021). Patients who did not opt out by a certain date would be automatically included. NHS intentions and transparency were further questioned as the consent was not well discussed and appeared only once in initial policy documents (Shaw and Erren, 2021). The term pseudonymisation (and its relationship to anonymisation) was confusing to many people outside academia. The following exchange during the oral evidence to the Health and Social Care Committee (2021) illustrates this:

"**Chair[10]:** (…) When that data is anonymised, so it cannot be traced back to an individual—

Phil Booth: Pseudonymised.

Barbara Keeley[11]: It is not anonymised, Chair.

---

[10] The chair of Health and Social Care Committee was Jeremy Hunt.

[11] Member of Parliament

> **Chair:** Let me finish my point. When it is anonymised—
>
> **Barbara Keeley:** But it is not anonymised, Jeremy.
>
> **Chair:** Barbara, can you let me finish my point? I am trying to establish what Phil believes so that I can understand it. If it is anonymised, so that it cannot be traced back to individual patients, are you objecting in those circumstances to that data being passed on to third parties for research purposes?
>
> ***Phil Booth*[12]*:*** We are objecting to what is done. Data is not anonymous. It is pseudonymised. The pseudonym is what allows it to be linked across an individual person's entire medical history. Medical history is a fingerprint unique to an individual, so pseudonymised data is all basically traceable back to an individual. If it was not, it would not be useful for research.

Chair: This is a very important point (...)".

Poor public engagement led to negative media coverage in the UK (Shaun, 2021; Craig, 2021), and medical associations (such as The British Medical Association and the Royal College of General Practitioners) called for more public engagement (BMA and RCGP, 2021). Because of the public controversy, one and a half million people in England initially opted out of the scheme (Wise, 2022). The government responded by extending the deadline for opting out and decided to engage the public further to establish public support and awareness of the scheme (Goldacre and Morley, 2022) and not proceed with the project until this has been achieved (NHS Digital, 2022). The involvement and engagement were ongoing at the time of writing this thesis. The involvement was especially targeting members of the public who remained sceptical to data linkage, for example, by including them as public contributors in the GP Data Patient and Public Engagement and Communications Advisory Panel (Phillips, 2022).

The cases of Care.data, Google Deep Mind and GPDPR show that transparency around big data projects is needed. There was a breach of the law in Google Deep Mind, but not in the other two cases. Therefore, these examples indicate that something more than following legal rules is needed to establish public support. This is called social licence and is discussed in the next section.

---

[12] Co-ordinator, from the medConfidential, the organisation in the UK arguing for confidentiality of patient medical data.

3.5. Social licence for big data research

The cases discussed in the previous section suggest that the public must have confidence that their data will be used only in an acceptable way (from their perspective) to support big data research (Taylor, 2014). Data reuse should not move beyond public expectations. Conformity with legal requirements was insufficient to gain public trust for Care.data and GPDPR. Taylor (2014) goes a step further and claims that the current data protection law is not good enough to deal with the challenges of big data research. Carter, Laurie and Dixon-Woods (2015) based on the experience of care.data project suggests the social licence as an answer to building public trust for big data research. They draw on non-medical literature to explain this term. The sociologist Everett Hughes (1958 as cited by Carter, Laurie and Dixon-Woods (2015)) argues in his work on the occupational relationship that in a relationship between a profession and the public, the latter grants the licence for that profession to conduct their work. Based on his research, the corporate social responsibility literature developed that concept. The public expects the corporation to follow societal standards and expectations that are beyond legal requirements. If the corporation breaks these rules, it might be punished by the public, for example, by boycotting its products and disturbing its activities, leading to financial loss (Gunningham, Kagan and Thornton, 2004). Therefore, the public grants a social licence to a group or organisation to complete certain activities (Paprica, De Melo and Schull, 2019).

The ‘moving beyond legal compliance’ principle would apply to big data research. There is no single definition of social licence to operate in non-health literature (Moffat *et al.*, 2016), and in health literature, the requirements for social licence still remain vague (Muller *et al.*, 2021). I follow the work of Carter, Laurie and Dixon-Woods (2015) as they propose three conditions for the social licence in big data research based on the public controversy around care.data. These are:

1) **Reciprocity**, there is a need for two-way communication. The public would improve their awareness of big data research and simultaneously, researchers could understand more public concerns and expectations. This principle aligns with public engagement in research.
2) **Non-exploitation**, the process should not disempower the public; this aligns with public involvement where members of the public could contribute to the process around governance for data linkage and the design of big data projects.
3) **Service of the public good**, the research should benefit the public**.** Public involvement and engagement could help researchers understand how the public perceives ‘good’.

If these three conditions are met, there is a high chance that the public will ‘grant’ a licence to operate for big data initiatives. There is already some indication that raising awareness of the challenges of big data research could improve public support (and thus social licence). For example, researchers could explain challenges around the time needed to obtain the opt-in consent and the risk of potential bias when not all groups are included. This could make members of the public more supportive of reusing their data (Hill *et al.*, 2013; Howe *et al.*, 2018). However, Hill *et al.* (2013), who explained that challenge to their participants, still found a minority of people would not accept even then the opt-out consent. Carter, Laurie and Dixon-Woods (2015) argue that it is impossible to convince everyone, but the social licence is about building a consensus with the broad public. However, this consensus is informal (Paprica, De Melo and Schull, 2019; Shaw, Sethi and Cassel, 2020) and offers flexibility to researchers regarding how they act (Ballantyne and Stewart, 2019).

Ongoing two-way communication with members of the public around big data research could allow further understanding of the public perspectives on big data research, as the public might not share policymakers’ priorities (Aitken et al., 2018). In their theoretical paper, Ford et al. (2019) argue that public involvement and engagement could bridge that gap. There is evidence in public involvement literature (outside big data) that involvement offers greater legitimacy for researchers (Manafo *et al.*, 2018). Public contributors could be a part of the process of creating research norms for big data research (Muller *et al.*, 2021). Research norms consist of governance and regulation that could guide research. These might not be popular among some academics, but they could secure a social licence for research (Dixon-Woods and Ashcroft, 2008). Aitken *et al.* (2019, p. 1), in their consensus statement on public involvement with big data research, go a step further and argue that “*the public should not be characterised as a problem to be overcome but a key part of the solution to establish beneficial data-intensive health research for all*.” There is emerging evidence that public contributors could be meaningfully involved in big data projects. For example, the SAIL Databank has a strong public and engagement component (Jones *et al.*, 2020b; Jones *et al.*, 2019; Jones, McNerney and Ford, 2014). The SAIL Databank is a Welsh initiative based at Swansea University and operates as a safe data haven that offers big data researchers access to routinely collected data. As a part of the involvement and engagement strategy the consumer panel (consisting of sixteen public contributors) provides advice on research proposals, discusses concerns around data protection, provides feedback on information shared with the general public and how to engage the public. Additionally, to that involvement, the SAIL Databank organises engagement activities (e.g. during science festivals) to reach a broader public. The case of the SAIL Database shows that public involvement in big data research could be a sustained organisational commitment and does not have to be tokenistic. However, there remain divided opinions among the public if public contributors should be involved in any governance. For example, Manhas *et al.* (2015, p. 95) found in their qualitative study with parents (who were lay people) that some participants felt that public contributors would not have the "*qualification to supervise that*".

Another aspect of social licence is transparency. A lack of transparency could make it challenging to secure public trust (Spencer *et al.*, 2016). The public has the right to be informed about the research progress (Shaw, Sethi and Cassel, 2020). Seldom-heard communities such as Black and South Asians especially want to receive more information about how their medical data is reused (Bailey Wilson *et al.*, 2022). This could involve providing information about potential data uses, benefits, risks, addressing safety concerns, sharing research results (Ballantyne and Stewart, 2019) and raising awareness of the private sector's role in healthcare (MORI, 2017). Public involvement and engagement could also assist in ensuring the transparency of big data initiatives. The growth of social media has the potential to facilitate that process (but there is also the possibility of inadvertently increasing barriers to contribution for not digitally active people). For example, in 2014, the University of Manchester's Health eResearch Centre launched the campaign Data Saves Lives. It aims to share the benefits of big data research and raise awareness among members of the public (Data Saves Lives, 2020). It is active on multiple social media channels (Twitter, Facebook and Instagram). Researchers and healthcare professionals were actively encouraged to use the hashtag on social media to share a positive message (Lawrence and Bradley, 2018). Hassan, Nenadic and Tully (2021) examined how the campaign took place on Twitter. Their findings show that the Data Saves Lives campaign successfully shared positive examples of big data research and offered Twitter users an opportunity to express their support for reusing health data for research. The case of Data Saves Lives shows that researchers can raise awareness with a broader public.[13] However, opponents of big data research could also utilise social media. During care.data controversy, both sides were active on Twitter (Hays and Daker-White, 2015).

This section has shown that public involvement and engagement could support establishing the social licence for big data research. Achieving a social licence is not a one-off event but rather an ongoing process (Muller *et al.*, 2021). Researchers need to continue the involvement and engagement activities throughout all stages of the research process. Therefore, it is necessary to understand how to involve and engage the public in big data research.

---

[13] For more on history and current development around Data Saves Lives campaign, please see chapter 9 at page 156

3.6. Conclusions

In this chapter, I discussed big data research. This aimed to set the context in which my thesis discusses public involvement and engagement. The chapter explored public views towards big data research initiatives and recognised that the public mostly supports it but also it has concerns, and these could be potentially visible among some seldom-heard communities (e.g. ethnic minorities). Thereafter, I discussed these concerns, focusing on lack of consent, confidentiality and data misuse. This was followed by an overview of three big data project initiatives in the UK that raised public controversy and the consequences of that. I concluded this chapter by showing that the literature suggests that public involvement and engagement could facilitate the establishment of a social license for big data research. However, no review has explored how and to what extent public involvement and engagement have been used to achieve that.

The previous chapter discussed public involvement and this one focused on big data research. As this thesis aims to explore how to involve and engage the public (especially seldom-heard communities) in big data research, it is necessary to discuss these two issues together, and this is the focus of the next two chapters. I decided to conduct a two-stage review. First, I did the scoping review on public involvement and engagement in big data research and then a systematic review of the delivery and effectiveness of strategies for involvement and engagement. The next chapter presents the methodology of this two-stage review, and the following one reports the findings.

4. Review protocol

4.1. Overview

The aim of this chapter is to start bringing together public involvement and big data research that have been discussed so far separately in previous chapters. It presents in verbatim the published paper in the BMJ Open titled *Public involvement and engagement in big data research: protocol for a scoping review and a systematic review of delivery and effectiveness of strategies for involvement and engagement* (hereafter "review protocol") (Teodorowski *et al.*, 2021a). The review protocol sets out the methods used in the review. It justifies decisions about the review question, inclusion criteria, search strategy, study selection, data extraction process, quality assessment, public involvement and dissemination plans. I briefly outline the review process and show how it builds on previous chapters.

The system logic model underpins the review to conceptualise the involvement and engagement around big data. It was developed based on the initial scope of the literature in public involvement and engagement, research team discussion and conversations with public contributors. The model explores the context, design of public involvement and engagement strategies, targeted population and outcomes. The idea was to develop the model as the review progressed.

The review protocol presents a two-step process. The first step was to conduct a scoping review to comprehensively map current evidence on public involvement and engagement in big data research. In the previous chapter on big data research, I have shown that public involvement and engagement could facilitate establishing a social licence for big data research, but there is no review exploring that potential. In the second stage, the studies discussing the delivery and effectiveness of involvement and engagement strategies were identified out of the papers included in the scoping review. This provided a further summary of evidence for researchers on how to involve and engage the public around big data research.

I involved two public contributors from the start, and they have been included in developing this protocol. The decision to involve them was based on evidence that public involvement could offer a lay perspective in the review. As the topic is public involvement, having their input would ensure that future findings would be of relevance to members of the public. I discuss the process and impact of involving two public contributors in chapter 8.[14]

---

[14] See more about public involvement in the review at page 146.

An open peer review process was used to review this paper. It means that both reviewers and I knew each other names, and the reports are published online (see publication history at BMJ Open (2021)).

Public involvement and engagement in big data research: protocol for a scoping review and a systematic review of delivery and effectiveness of strategies for involvement and engagement.

4.2. Abstract

Introduction: Big data research has grown considerably over the last two decades. This presents new ethical challenges around consent, data storage and anonymisation. Big data research projects require public support to succeed and it has been argued that one way to achieve this is through public involvement and engagement. To better understand the role public involvement and engagement can play in big data research, we will review the current literature. This protocol describes the planned review methods.

Methods and analysis: Our review will be conducted in two stages. In the first stage, we will conduct a scoping review using Arksey and O'Malley (2005) methodology to comprehensively map current evidence on public involvement and engagement in big data research. Databases (CINAHL, Health Research Premium Collection, PubMed, Scopus, Web of Science) and grey literature will be searched for eligible papers. We provide a narrative description of the results based on a thematic analysis. In the second stage, out of papers found in the scoping review which discuss involvement and engagement strategies, we will conduct a systematic review following PRISMA guidelines, exploring the delivery and effectiveness of these strategies. We will conduct a qualitative synthesis (Thomas and Harden, 2008). Relevant results from the quantitative studies will be extracted and placed under qualitative themes. Individual studies will be appraised through MMAT (Hong et al., 2018), we will then assess the overall confidence in each finding through GRADE-CERQual (Lewin et al., 2015). Results will be reported in a thematic and narrative way.

Ethics and dissemination: This protocol sets out how the review will be conducted to ensure rigour and transparency. Public advisors were involved in its development. Ethics approval is not required. Review findings will be presented at conferences and published in peer-reviewed journals.

Keywords: Big data, PPI, public involvement, patient engagement, consumer participation, governance

Strengths and limitations of this study

- This is the first review exploring public involvement and engagement in big data research
- The search is limited to studies published in English.
- Lack of clarity and consistency with the use of the terms public involvement, engagement, and big data could impact our search results. However, we will undertake additional searching techniques to mitigate this limitation.

4.3. Introduction

What is the problem?

Over the last two decades, the ongoing digitalisation of information has allowed the creation and linkage of large, multi-source health data sets to provide novel healthcare applications. This is often called ‘big data’, but the concept itself is unclear and heavily debated (Mehta and Pandit, 2018). However, this growing area of research has the following characteristics: large volume, high velocity, huge variety, veracity and value (ibid). Multiple stakeholders use big data for research; clinical management; audit; service evaluation, or statistical purposes. The UK has been a global leader in big data research. Large projects include, at national level, OpenSAFELY (OpenSAFELY, 2021) and regionally located projects such as Children Growing Up in Liverpool (C-GULL) (University of Liverpool, 2020) (to name a few). The overriding aims of big data research projects are to deliver more efficient healthcare (Raghupathi and Raghupathi, 2014), and to reduce health inequalities (Zhang *et al.*, 2017).

The use of big data for research presents ethical challenges (Lipworth *et al.*, 2017). Traditionally, a person consents to participate in a research study, whereas when large quantities of data are collected, it is not often apparent how it will be (re)used in the future. Data can be collected for one purpose (e.g. audit or to collect groups statistics) and only later shared or linked for research. Secondly, even when big data is anonymised, in theory, individuals can be still re-identified (Lipworth *et al.*, 2017). Thirdly, digitalised data needs to be stored - sometimes in various places and hosted by both public institutions and private companies. Despite these ethical issues (consent, anonymisation, data storage and access), the literature shows that the public mostly supports big data usage in research (Aitken *et al.*, 2016), but is sceptical toward current governance mechanisms (Aitken, Cunningham-Burley and Pagliari, 2016) and concerned about associated risks such as breach of privacy, generating waste of unused information and usage of data for profit rather than for the public good. Big data is still new, and thus it often outpaces governance structures and regulation. Even if researchers meet the legal requirements, the public might not be supportive of their actions (Carter, Laurie and Dixon-Woods, 2015; Aitken *et al.*, 2019). Controversial cases can undermine public trust in big data. For example, the case of DeepMind in the UK illustrated these dangers: the NHS breached data protection legislation by sharing patients' data (without properly informing them) with the Google-owned company (Revell, 2017). Low public engagement and lack of transparency in the care.data project in the UK (Taylor, 2014) led to its eventual closure. The public might perceive the risk-benefit ratio as unfavourable for them and therefore not want to support or participate in the research. Also, it could foster general distrust in healthcare professionals.

What is the solution?

The concept of trust is vital in building a positive relationship between researchers and the public (Aitken, Cunningham-Burley and Pagliari, 2016). Improving people's knowledge, through public engagement, of how big data research works can improve public support for using health data (Hill *et al.*, 2013). For example, the #DataSavesLives initiative raises awareness of the benefits of health data research to gain public trust (Data Saves Lives, 2020). Secondly, researchers should involve the public in developing transparent, accountable policies and governance processes (Ford *et al.*, 2019). Public involvement and engagement are crucial mechanisms to develop governance policies and build trust between the public and researchers. Public involvement should be genuine. It should not be carried out with the sole aim of benefiting researchers; be tokenistic or mislead the public (Beier, Schweda and Schicktanz, 2019). Extensive evidence shows that successful public involvement can lead to service improvement (Lalani *et al.*, 2019; Ocloo and Matthews, 2016; Crawford, 2002; Mockford *et al.*, 2012), raises awareness of services (Mockford *et al.*, 2012), and brings together patients' and researchers' priorities (Sofolahan-Oladeinde *et al.*, 2017).

Public involvement in big data research has context-related challenges. In traditional research, a participant and a researcher would have some contact. In contrast, big data research includes large groups of people (who might not necessarily be aware that a particular research team uses their data), thus creating a feeling of remoteness between researchers and the public (Aitken *et al.*, 2019). Therefore, building trust between the public and researchers is more challenging. Transparent governance policies need to be developed with public involvement to ensure transparency. Lay people can be members of ethics and governance committees overseeing research projects, ensuring public voices are heard. Researchers need big data specific recommendations on involving and engaging the public. However, the literature on public involvement and engagement in big data research is still limited.

Why is this review needed?

Systematic and narrative reviews that have explored the public attitudes towards big data have typically focused on trust or attitudes towards using big data for research (Aitken *et al.*, 2016; Stockdale, Cassell and Ford, 2019; Kalkman *et al.*, 2019b; Howe *et al.*, 2018). However, how and to what extent public involvement and engagement is used in establishing trust for big data research (e.g., organising and maintaining large health data sets and its governance policies) has received less attention. To our knowledge, there is no review covering our objectives published or registered on Prospero or Cochrane databases.

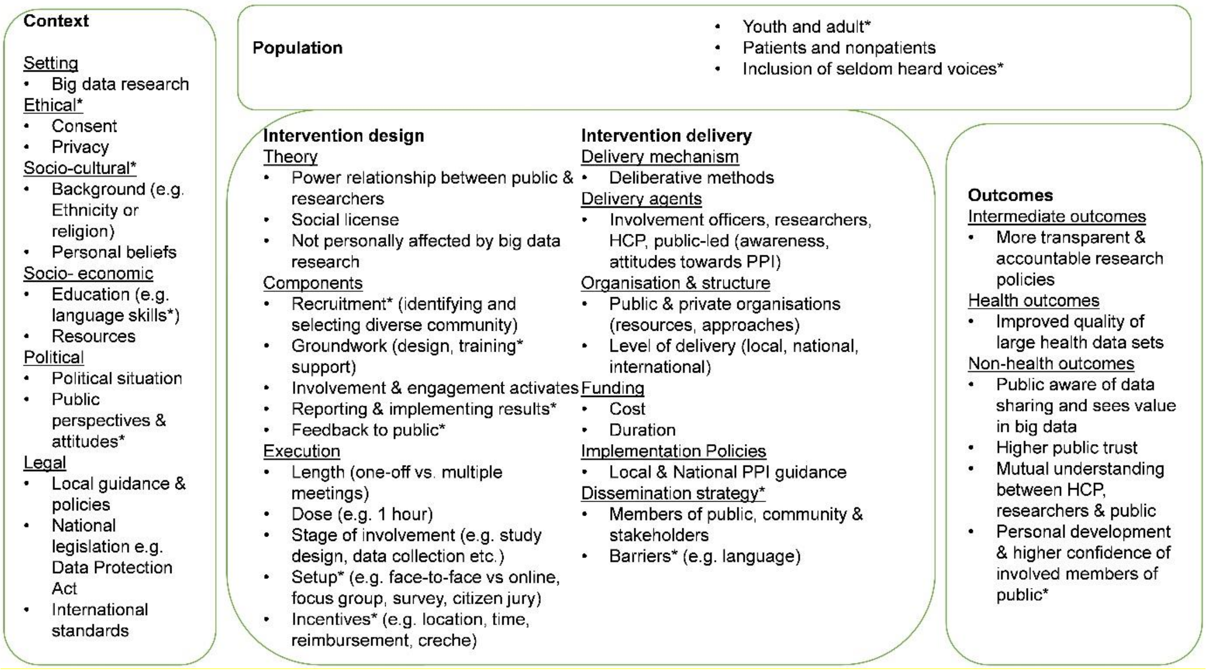


Figure 1 System logic model of public involvement and engagement in big data research.

To better understand the complexity of public involvement and engagement in big data research, we developed a system logic model (see Figure 1) following Rohwer et al. (2017) guidance. Graphic presentations (such as logic models) can be used in reviews to identify relevant elements and the relationships between them. This model is based on team discussion, a preliminary scoping of literature, and public advisors' feedback. We used asterixis (*) to record those sections which were suggested by public advisors. Our model puts special emphasis on four related sections: context, design of public involvement and engagement strategies, targeted population, and outcomes. As our review progresses, we will develop the logic model, and present the final version in the report of our review's findings. We hope that the model will assist in interpreting the findings and identifying gaps in the literature.

4.4. Review objectives

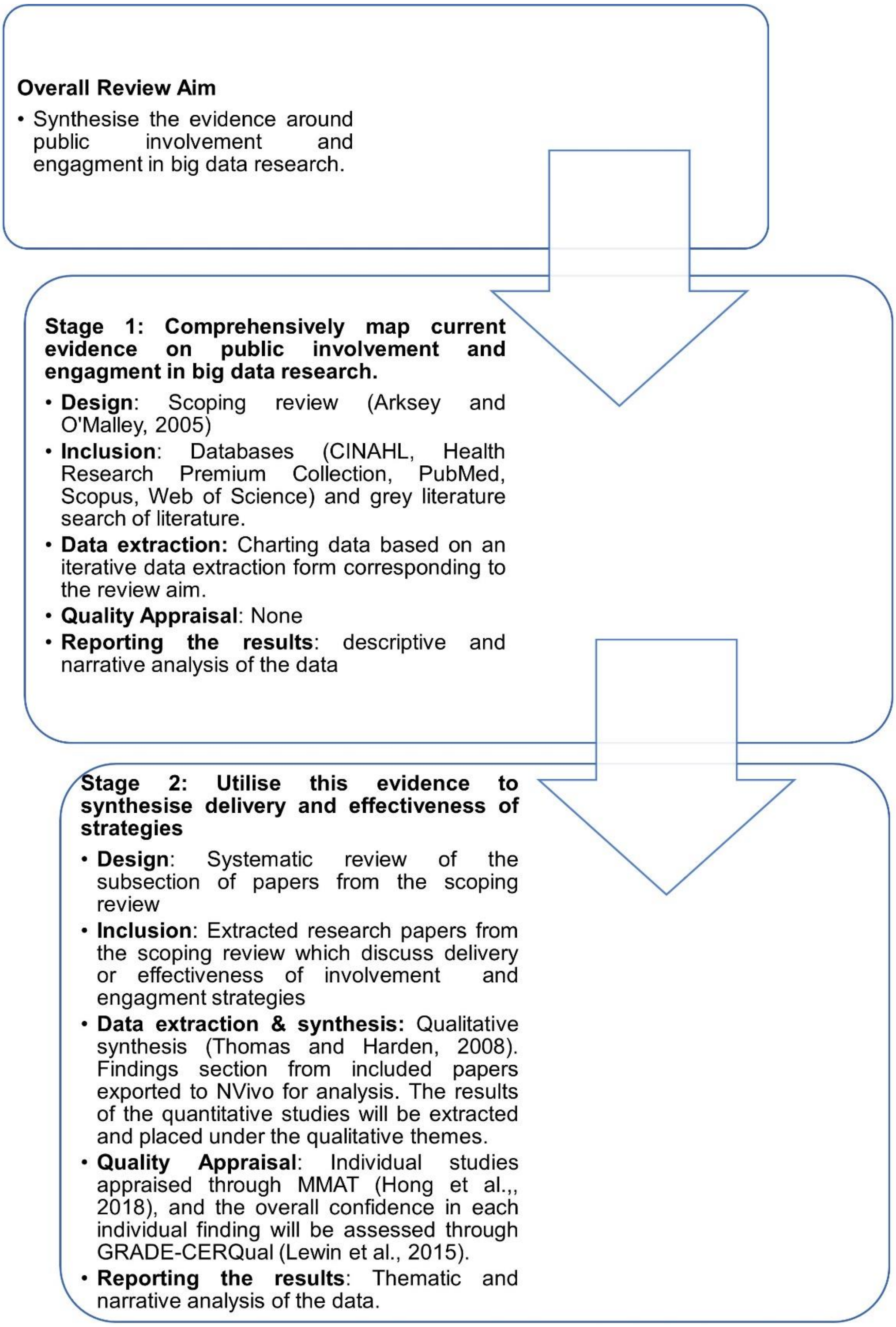
The purpose of this review is to synthesise the evidence on public involvement and engagement in big data research. We have two objectives:

1. Comprehensively map current evidence on public involvement and engagement in big data research (scoping review).
2. Utilise this to synthesise evidence on the delivery and effectiveness of involvement and engagement strategies (systematic review).

4.5. Methods and analysis

4.5.1. Design

The review will be conducted in two stages as illustrated in Figure 2 (Gough, Thomas and Oliver, 2012). These stages will complement each other and assist in flexibly understanding the phenomenon. Firstly, the literature on public involvement and engagement in big data research will be explored by conducting a scoping review. We follow Arksey and O'Malley (2005) framework and its further iterations (Colquhoun *et al.*, 2014; Levac, Colquhoun and O'Brien, 2010). The scoping review will allow us to clarify concepts, illustrate current evidence in the field and gaps in research (Munn *et al.*, 2018). In the second stage, out of papers identified in the scoping review, we will extract those discussing involvement and engagement strategies to explore their delivery and effectiveness. The findings from the systematic review will inform researchers on best practice and identify any conflicting views (Munn *et al.*, 2018). To further enhance the quality of this review, we follow PRISMA reporting guidelines (Moher *et al.*, 2015).



Overall Review Aim

- Synthesise the evidence around public involvement and engagment in big data research.

Stage 1: Comprehensively map current evidence on public involvement and engagment in big data research.

- **Design**: Scoping review (Arksey and O'Malley, 2005)
- **Inclusion**: Databases (CINAHL, Health Research Premium Collection, PubMed, Scopus, Web of Science) and grey literature search of literature.
- **Data extraction:** Charting data based on an iterative data extraction form corresponding to the review aim.
- **Quality Appraisal**: None
- **Reporting the results**: descriptive and narrative analysis of the data

Stage 2: Utilise this evidence to synthesise delivery and effectiveness of strategies

- **Design**: Systematic review of the subsection of papers from the scoping review
- **Inclusion**: Extracted research papers from the scoping review which discuss delivery or effectiveness of involvement and engagment strategies
- **Data extraction & synthesis:** Qualitative synthesis (Thomas and Harden, 2008). Findings section from included papers exported to NVivo for analysis. The results of the quantitative studies will be extracted and placed under the qualitative themes.
- **Quality Appraisal**: Individual studies appraised through MMAT (Hong et al.,, 2018), and the overall confidence in each individual finding will be assessed through GRADE-CERQual (Lewin et al., 2015).
- **Reporting the results**: Thematic and narrative analysis of the data.

Figure 2 Systematic map of the review process

4.5.2. Stage 1: Scoping review

4.5.2.1. Search strategy

We will search the following databases CINAHL, Health Research Premium Collection, PubMed, Scopus, Web of Science and check sources of grey literature related to public involvement such as the Patient-Centred Outcome Research Institute. The first hundred hits (to be inclusive but practical) of Google Scholar search results will be scanned for inclusion. We will also hand-search papers in the journals Health Expectations, BMC Research Involvement and Engagement, and the International Journal of Population Data Science. This will be followed by snowball sampling where we will check references in included papers to identify additional studies for inclusion and consult with experts about relevant papers. Big data research is a newly developing field; for instance, MeSH terms ‘big data’ was added in 2019. Thus, to capture these recent developments, we will restrict searches to a start date of 2010 and will update our searches prior to the final submission of our findings.

We developed the search strategy in partnership with an information specialist and tested this through an iterative process. It consists of both Boolean operators and where possible MeSH (PubMed) or subject heading (CINAHL). Three databases were searched in a test run and yielded a large number of references that were not relevant to our review aims. Therefore, we decided to include the further term “data governance” as we expect that most of the public involvement and engagement in big data research would be at the stage of developing and maintaining data sets. The summary of the search strategy is presented in Table 1.

Public	“advisory group” OR carer* OR citizen* OR client* OR communit* OR consumer* OR famil* OR lay OR nonpatient* OR participant* OR patient* OR public OR relative* OR representative* OR stakeholder* OR “steering group*” OR survivor* OR user*
Involvement or engagement	advocacy OR collaborat* OR co*production OR consult* OR empower* OR engage* evaluat* OR involv* OR particip* OR partner* OR PPI OR organi*ation* OR representation*
Big data	database OR “big data" OR “data science” OR “data mining” OR “datasets” OR “data analytics” OR “data sets”
Public Involvement	“patient participation” OR “consumer participation” OR “client participation” OR “community participation”

Data governance (only Health Research Premium Collection, Scopus & Web of Science)	“data governance”

Table 1 Search strategy

4.5.2.2. Inclusion & exclusion criteria in the scoping review

Public involvement and engagement can take place at any stage of a big data research project. Thus, we will include papers relating to any public role or contribution to big data research. These roles can include permission to use data, involvement in defining aims or design, and participation in decision-making processes (also the public may become members of a research team) (Beier, Schweda and Schicktanz, 2019).

Previous reviews (Dawson *et al.*, 2018; Harrison *et al.*, 2019b; Lalani *et al.*, 2019) have noted that a lack of one generally accepted definition of public involvement makes searching databases challenging. Hence, the definition of public involvement and engagement in the literature lacks consistency (Mockford *et al.*, 2012). Involvement, engagement, participation are often used interchangeably in the literature but do not necessarily have the same meaning (Islam and Small, 2020). We follow the INVOLVE (2020) definition of public involvement and engagment:

Public involvement – “*research being carried out ‘with’ or ‘by’ members of the public rather than ‘to’, ‘about’ or ‘for’ them.*”

Consultation – researchers discussed the project with members of the public. It was more of ‘to’, ‘about’ or ‘for’ rather than ‘with’ or ‘by’ them.

Public engagement – “*information and knowledge about research is provided and disseminated.*” – this usually takes place after the project is concluded.

INVOLVE’s definition of involvement sees an equal relationship between researchers and the public. Thus, involvement should mean co-design and co-production rather than just consultation. However, we will not exclude papers that do not meet this requirement but note it. Thus, included papers will be assigned one of three named categories: green (when it meets the definition of public involvement), blue (when consultation took place) and amber (where only the engagement occurred).

Multiple definitions of big data exist (Mehta and Pandit, 2018). To broadly map the current evidence, we use a definition which focuses on big data in the healthcare setting:

Big data – data which is challenging to manage through traditional analytic tools and meets the 5V characteristics: volume, velocity, variety, veracity and value (Mehta and Pandit, 2018).

The volume suggests that there may be a high quantity of data available potentially on millions of patients. The variety means heterogeneity of data collected as it can come in various formats (e.g. images, text). The velocity means that it can be collected swiftly from various sources. Veracity relates to the accuracy and identification of any biases. The value refers to the ability of results from research based on big data to guide decisions. Big data sources can be internal (e.g. patients record, healthcare professional notes, generated through apps or social media) and external (e.g. private companies or governmental institutions).

To map a range of studies, we will keep the selection criteria purposefully broad. Papers can discuss single research project or data sharing initiative. All study designs will be included. Papers can be (but not limited to) original research, an evaluation, a review, an expert opinion, or a commentary that explores any public involvement and engagement in big data research.

We will exclude a paper if it:

- does not discuss public involvement or engagement
- does not discuss a patient-related (or health-related) application
- the full text is not available in English

4.5.2.3. Study selection

Prior to the screening stage, we will organise a meeting for everyone involved in study selection process during which we will jointly scan a sample of 100 papers. We will record and discuss our disagreements. Then separately, we will scan all papers' eligibility, based on the title and then the abstract identified in the database searches. At each stage, two reviewers will be involved. The first reviewer will scan all papers and the second will check a random sample (20% of all papers). Reasons for exclusion will be recorded. If there are any disagreements, we will include a third reviewer. Then the full text will be screened by two reviewers, checking if the paper meets the inclusion and exclusion criteria. We will meet after each screening stage (title, abstract and full paper) to discuss ours experiences.

4.5.2.4. Data extraction

We will use an extraction form which will cover the following information:

- Paper aim
- Design
- Country
- Demographics of participants (also record if there are a seldom-heard group)
- Context
- Process of involvement or engagement

- Funding
- Legal or ethical issues
- References to guidance & policies
- Challenges and facilitators of public involvement and engagement

We see the extraction stage as an iterative process. After extracting initial papers, we will discuss if the extraction form is applicable in our review during team meetings. Where necessary, we will revise it. Each paper will be extracted by one reviewer and the second will validate data extraction.

4.5.2.5. Reporting the results

We will provide a descriptive and narrative analysis of the data. These will be used to develop the system model. Then, we will discuss the implication of the findings for researchers and policy.

4.5.3. Stage 2: Systematic review

4.5.3.1. Criteria for inclusion

Out of papers identified in the scoping review, we will extract qualitative and quantitative studies that discuss the delivery or effectiveness of involvement and engagement strategies.

4.5.3.2. Data extraction and synthesis

We will follow Thomas and Harden (2008) stages of qualitative synthesis. We plan to extract all findings sections from included papers and upload them to NVivo for analysis. Coding will be done inductively to develop descriptive themes to further our review aims and develop the system model. Thus, we want to ensure that no prior framework will influence us in identifying the relevant evidence. The relevant results from the quantitative studies will be extracted and placed under qualitative themes, as we do not expect that meta-analysis will be possible. At the last stage of the synthesis, we go beyond the descriptive themes and analyse them in the context of the aims of our review. The results will be provided in a thematic, narrative way and utilised to develop the system model.

4.5.3.3. Studies and Findings Appraisal

Using MMAT(Hong *et al.*, 2018) we will systematically appraise all studies included in the systematic review. However, no paper will be excluded if it scored low. The overall confidence in each individual paper's qualitative findings will be assessed through GRADE-CERQual (Lewin *et al.*, 2015). We will not assess the overall confidence in quantitative studies as these will be placed under the qualitative themes. This will allow researchers to make judgments about the quality of available evidence.

4.6. Patient and public involvement

Stakeholders (including patients and health professionals) can be involved in systematic reviews (Cottrell *et al.*, 2015; INVOLVE, 2012b). They can enhance the quality of the review by advising on the review questions and its scope. This ensures transparency and accountability, especially if the review aims to shape practice and improves relevance to those who this review seeks to influence (e.g. practitioners and public). Similarly, for scoping reviews Arksey (2005) recommends, and Levac et al. (2010) argue that consultation is a part of the review process. We have involved two public advisors who assisted in designing this protocol and will be co-authors on all publications. They have experience of conducting systematic reviews, represent seldom-heard communities and SA is a Big Data Ambassador for Care and Health Informatics theme within ARC NWC (Applied Research Collaboration North West Coast). They will be involved in the whole review process, with a particular emphasis on interpreting the findings and developing recommendations for both research and practice. We will report on public involvement using the GRIPP2 checklist (Staniszewska *et al.*, 2017a).

4.7. Limitations

The main limitation of our review is the exclusion of non-English papers. There is a possibility that some papers relevant to our review aims will be excluded and this will impact our findings. Secondly, as already mentioned the lack of clear definitions of public involvement. engagement and big data make any search strategy challenging, and potentially some relevant papers might not be included. However, we will undertake all reasonable steps to balance this limitation by involving experts and checking references in included papers.

4.8. Ethics and dissemination

We have published this protocol and engaged with public advisors to ensure transparency and rigour of our review process. As we are using already published data, there is no need to apply for ethical approval to conduct our study. We will present our findings at relevant conferences and publish in a peer-reviewed journal.

4.9. Conclusions

This review will synthesise the current literature on public involvement and engagement in big data research. Our work is timely as it is expected that big data research in healthcare will continue to grow rapidly. There will be increasing interest in developing large health data sets by researchers, funders, and governmental bodies. Previous research shows the need for synthesising the current evidence. Mouton et al. (2018) discussed issues around patient trust and big data, and how they viewed healthcare practitioners and professionals' involvement in funding or controlling big data research. They believed that patients were not interested or did not understand big data – and therefore, should not be involved in its governance. Their comments included remarks that patient groups are not important and the belief that patients' involvement in governance would be pointless. On the other hand, Aitken, Cunningham-Burley and Pagliari (2016) explored the similar issues with members of the public who presented opposite views on lay involvement in data governance. Participants believed that members of the public could promote accountability of big data research. Public involvement has the potential to shift perspectives and bridge the gap between researchers and the public, and help the development of big data research that has wider spread public support and buy-in.

5. Review findings

5.1. Introduction

This chapter brings together public involvement and big data by presenting the findings from the two-stage review using the protocol that has been presented in the previous chapter. This consists of the scoping review on involvement and engagement in big data research and then the systematic review of delivery and effectiveness of strategies for involvement and engagement. First, I discuss the methods used in the reviews and reflect on how the protocol has been followed. Second, I present the review findings. Thereafter, I reflect on the importance of these findings and identify gaps in the literature. Then, using review findings, I test and develop the system logic model for public involvement and engagement in big data research. Finally, I briefly discuss the review limitations.

5.2. Methods

5.2.1. Searches

Following the developed search strategy,[15] databases CINAHL, Health Research Premium Collection, PubMed, Scopus and Web of Science were searched for papers in September 2021. The search covered papers published after 2010. Additional manual searches took place. I screened the first 100 hits in Google search, journals (BMC Research Involvement and Engagement, International Journal of Population Data Science and Health Expectations), grey literature (Patient Outcome Research Institute database first 100 hits were screened), and shared the call for inclusion on Twitter to reach experts in the field. In contrast to the initial plan, the references of included papers were not screened for potential inclusion. After speaking with the rest of the research team, it was considered impractical due to the number of included papers.

5.2.2. Screening

I took the lead in the screening process by screening all papers. Two public contributors[16] and another doctoral student supported me by screening 20% of papers at each stage (title, abstract and full paper). We met regularly as a group to discuss any discrepancies.

5.2.3. Data extraction and analysis

I developed the initial data extraction form, and we piloted it with the team, made up of my supervisors, public contributors and another doctoral student. The final data extraction form is available in Appendix 1. I extracted data from all papers in the first instance. Then, two public

---

[15] See Table 1 at page 65

[16] See more about public involvement in the review at page 146.

contributors, another doctoral student and two supervisors double-checked each included paper. I then organised the extracted data to present in a descriptive and narrative way.

5.3. Stage 1: Scoping Review

The searches produced 4054 papers. Manual searches added additional 11 papers. After the removal of duplicates, 3540 articles were screened for inclusion in the review. 3342 papers were excluded based on the title and abstract. The full-text screen took place for 198 papers, and 53 were included in the review. Figure 3 shows the PRISMA flow chart of the screening process.

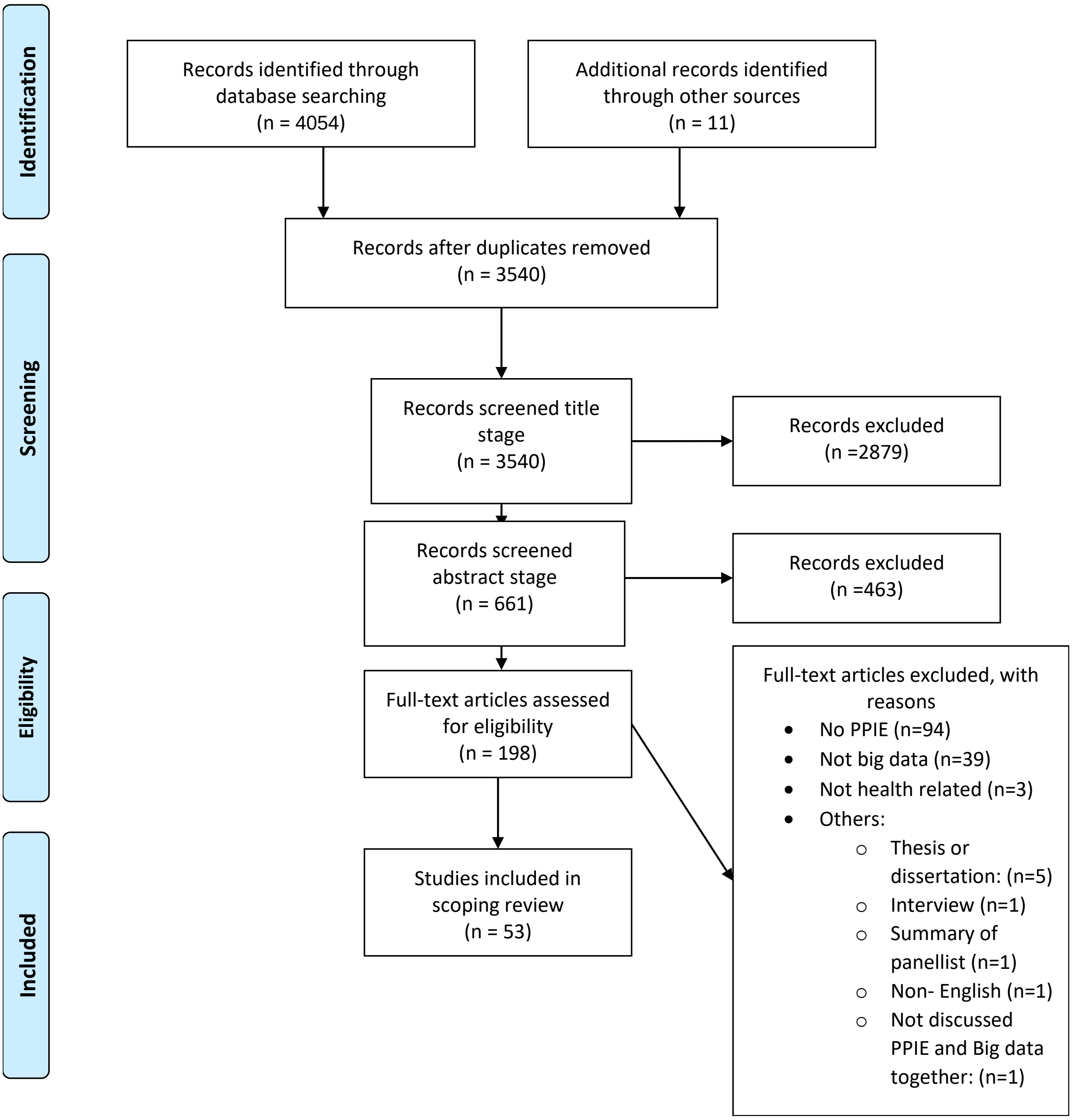


Figure 3 PRISMA flow chart as adopted by Moher *et al.* (2009) and Moher *et al.* (2015)

5.3.1. Study Characteristics

The most popular study design was discussion (n=28), followed by review (n=5), qualitative (n=5), opinion, letter, commentary or editorial (n=4), evaluation (n=3), protocol (n=2), ethnographic or descriptive case study (n=2), public deliberations (n=1), action research (n=1), quantitative (n=1), mixed method (n=1). The papers were around issues in the UK (n=19), US (n=10), Canada (n=7), New Zealand (n=3), Netherlands (n=1), Portugal (n=1), France (n=1), South Africa (n=1), Australia (n=1), Germany (n=1), and Africa (n=1). Twelve papers did not specify a geographical location, and some included more than one. The most popular activities with the public (following NIHR definitions) were involvement (n=45), then engagement (n=25) and the least consultation (n=7). Some papers have discussed more than one type of activity. The suggested demographics of communities to be involved and engaged were diverse. These included patients (including consumers and service users) (n=12), affected or vulnerable groups (n=8), indigenous communities (n=6), focusing on specific characteristics (e.g. gender, age, income, education or geography) (n=5), careers (n=4), the general public (n=3), ethnic minorities (n=3), patient representative or community leaders (n=3) and study participants (n=1). Table 2 presents the characteristics of included papers.

Author (Year)	Design	Location	Demographics to involve & engage	Types of activities	Area of interest
Baart & Abma 2010	Action Research	Netherlands		Involvement, Engagement	Involvement in psychiatric genomics research.
Ballantyne and Style, 2017	Discussion	New Zealand	Lay, gender and Maori representation.	Involvement, Engagement	Expert health-data research ethics committee.
Ballantyne and Stewart, 2019	Discussion	UK	Affected group. Priority is given to vulnerable patient groups.	Involvement, Engagement	Public and private sectors collaboration to share, analyse and use biomedical big data.
Beyer et al., 2010	Qualitative	US	Caucasian, Hispanic, or Taidam/Lao. Represented various education, income, and other characteristics.	Involvement, Consultation	Geocoded health information and experiential geographical information in a GIS environment.
Bharti et al., 2021	Discussion	UK		Engagement	Securing public trust and the importance of public engagement.
Bot et al., 2019	Discussion	US	Under-represented populations.	Involvement	Decentralisation of governance.
Coulter 2021	Editorial	UK	General public.	Involvement	NHS Digital plans to update its systems from patient data from GP records.

Dankar, et al., 2018	Discussion			Engagement	Data governance in population genome projects.
de Freitas et al., 2021	Protocol	Portugal	Patients and informal carers.	Involvement	Co-production of a people-centred model for the public in decision-making processes about data reuse.
Deverka et al., 2019	Public deliberations	US	Diverse geographic and individuals with chronic illness.	Involvement, Consultation	Recommendations for medical information commons design and management.
Duchange et al., 2014	Discussion	France (EU project)	Representatives of patient organisations.	Involvement, Engagement, Consultation	Ethics committee.
Erikainen et al., 2020	Qualitative	UK		Involvement	Governance of population-level biomedical research.
Evans et al., 2020	Qualitative	US	Individuals with OUD and their families.	Involvement, Engagement	Reuse of big data on opioid use.
Fernando et al., 2019	Letter	South Africa	Traditional community leaders.	Involvement, Consultation	Data governance model in biobanking and data sharing.
Fleurence et al., 2014	Discussion	US	Patients.	Involvement	National research network (PCORnet).
Funnell, et al., 2020	Discussion	Canada	Indigenous communities.	Involvement	Community-based participatory research (CBPR) methods in a project using previously collected data to examine end-of-life health care.
Gallier et al., 2021	Discussion	UK		Involvement, Engagement	PIONEER infrastructure and data access processes.
Goytia et al., 2018	Qualitative	US	Patients.	Involvement, Engagement	Views on big data research.
Henare at al., 2019	Opinion	New Zealand	Indigenous people.	Involvement, Engagement	Roadmap for neuroendocrine tumour research to reflect the values of Indigenous people.

Hudson et al., 2020	Discussion		Indigenous population.	Involvement	Indigenous communities' views around the sharing of genomic data.
Hurt et al., 2019	Discussion	UK		Involvement, Engagement	Design of HealthWise Wales.
Jewell et al., 2019	Evaluation	UK	Service users and carers.	Involvement	Advisory group.
Jones et al., 2014	Evaluation	UK	Consumers. At least one representative of an ethnic minority.	Involvement	Consumer Panel.
Jones et al., 2019	Discussion	UK		Involvement, Engagement	SAIL Databank.
Jones et al., 2020	Evaluation	UK	Inclusive of all ages, ethnic groups, cultures, socioeconomic levels, lifestyles and other definable interests.	Involvement Engagement,	SAIL Databank and related population data science initiatives.
Kalkman et al., 2019	Systematic review			Involvement, Engagement	Ethical guidelines for principles and norms pertaining to data sharing.
Kirkham et al., 2021	Qualitative		People with lived experience of mental illness and experience with data science or research methods.	Involvement	Best practice checklist for use in mental health data science.
Luna Puerta et al., 2020	Scoping review			Involvement	Reporting impact of public involvement in biobanks.
Manrique de Lara and Peláez-Ballest	Narrative review			Involvement, Engagement	Bioethical perspectives of big data

as, 2020					
Milne et al., 2021	Discussion	US and North America		Involvement	Data trust model in the governance of biobanks.
Milne & Brayne, 2020	Discussion			Involvement	Data governance in dementia.
Mourby et al., 2019	Discussion	UK		Involvement, Engagement	Obstacles preventing data linkage research from reaching its full potential.
Murtagh et al., 2018	Ethnographic case study	UK	Participants of genomic studies.	Involvement, Engagement	Foundational principles of data sharing infrastructure.
Nelson and Burns, 2020	Discussion	UK	Most affected communities by the research.	Engagement	ADRC NI approach to public engagement.
Newburn et al., 2020	Discussion	UK	Service users. One activity targeted ethnic minorities.	Involvement, Engagement	Service user participation in a data linkage study.
Nunn et al., 2021	Mixed method	Australia		Involvement	Involvement in genomic research.
O'Doherty et al., 2011	Discussion	Canada	Historically disadvantaged groups.	Involvement, Engagement	Biobank governance and principles to form governance structures.
O'Doherty et al., 2021	Commentary			Involvement	Functions of good governance.
Ohno-Machado et al., 2014	Discussion	US	Patients.	Involvement, Consultation	Setting up of the pSCANNER.
Omar et al., 2020	Discussion			Involvement, Engagement, Consultation	European network of excellence for big data in prostate cancer.
Paprica et al., 2020	Discussion	Canada	Communities facing long-standing inequalities that are affected by the research.	Involvement, Engagement	Establishment and operation of data trusts.

Patel et al., 2021	Quantitative	UK		Involvement	The use of remote consultation and prescribing of psychiatric medications.
Pavlenko et al., 2020	Systematic review			Involvement	Governance in clinical data warehouses internationally.
Rowe et al., 2021	Discussion	Canada, New Zealand and US	Indigenous people.	Involvement	Principles for linking indigenous population data.
Shaw et al., 2020	Discussion	US, Canada, UK	General public, specific communities (e.g. African Americans, indigenous people, people with disabilities, people living with homelessness).	Engagement	Social license for big data initiatives.
Sleigh and Vayena, 2021	Descriptive case study	Germany, UK.	General public.	Engagement	Visual public engagement campaigns.
Teng et al., 2019	Discussion	Canada		Involvement	Public deliberation event on the data linkage and reuse for research.
Tindana et al., 2015	Review	Africa	People affected by the research.	Involvement, Engagement, Consultation	Community engagement in biomedical and genomic research.
Townson et al., 2020	Discussion	UK		Involvement, Engagement	A model of public involvement and engagement.
Vayena & Blasimme, 2017	Discussion		Patients.	Involvement	Models of informational control in data-intense healthcare and clinical research.
Weich et al., 2018	Protocol	UK	Mental health users and carers, people with	Involvement	Spatial and temporal variation in the use, effectiveness and cost of CTOs through the

			lived experiences. Ensure diversity of age, gender and ethnicity.		analysis of routine administrative data.
Willison et al., 2019	Discussion	Canada	Patients representative (with diabetes or Seldom-heard: (francophone, immigrant, indigenous).	Involvement	Governance model for health data repositories.
Xafis and Labude, 2019)	Discussion			Involvement, Engagement	Ethics framework for big data in health and research.

Table 2 Summary of included papers in the scoping review

5.3.2. Activities

5.3.2.1. Involvement

Public contributors have various roles in big data research. First, they could contribute to a specific research project. Thus, public contributors are involved at all stages, from the study design, through identifying research questions, analysis to dissemination (Evans *et al.*, 2020; Manrique de Lara and Peláez-Ballestas, 2020; Jewell *et al.*, 2019; Xafis and Labude, 2019; Newburn *et al.*, 2020; Goytia *et al.*, 2018; Weich *et al.*, 2018; Fernando, King and Sumathipala, 2019; Tindana *et al.*, 2015; de Freitas *et al.*, 2021; Funnell *et al.*, 2020). Public contributors could act as co-investigators in these big data research projects (Newburn *et al.*, 2020).

The other role is around data governance. Public contributors (or representatives of patients organisations) could be meaningfully involved in (joint) data governance to ensure that research is done ethically (in terms of public interest and sensitivity risk), for example, by advising, co-finding new solutions or co-creating guidance and policy (Ballantyne and Style, 2017; Hudson *et al.*, 2020; Omar Muhammad *et al.*, 2020; Gallier *et al.*, 2021; Ballantyne and Stewart, 2019; Fleurence *et al.*, 2014; Hurt *et al.*, 2019; Luna Puerta *et al.*, 2020; Mourby *et al.*, 2019; Ohno-Machado *et al.*, 2014; Pavlenko, Strech and Langhof, 2020; Milne and Brayne, 2020; Bot, Wilbanks and Mangravite, 2019; Kalkman *et al.*, 2019a; Milne, Sorbie and Dixon-Woods, 2021; O'Doherty *et al.*, 2021; Henare *et al.*, 2019; Jones *et al.*, 2020b; O'Doherty *et al.*, 2011; Paprica *et al.*, 2020; Jones, McNerney and Ford, 2014; Murtagh *et al.*, 2018; Patel *et al.*, 2021; Willison *et al.*, 2019). Working with the public offers a lay perspective and could ensure that data access and research are in the public interest and, thus, establish public trust (Ballantyne and Style, 2017; Gallier *et al.*, 2021; Hurt *et al.*, 2019; Jones *et al.*, 2019; Mourby *et al.*, 2019; Milne, Sorbie and Dixon-Woods, 2021; Jones, McNerney and Ford, 2014). Governance groups could be chaired or co-chaired by a public contributor, and the majority of members of these groups might be members of the public (Hurt *et al.*, 2019; Ohno-Machado *et al.*, 2014; Milne, Sorbie and Dixon-Woods, 2021; Willison *et al.*, 2019). If there is more than one governance group in the organisation, public contributors could sit on different panels (Jones *et al.*, 2019; Ohno-Machado *et al.*, 2014; Jones *et al.*, 2020b; Jones, McNerney and Ford, 2014).

Public contributors who are members of governance bodies act as the main big data advocates (Jones *et al.*, 2020b). However, their voice should be of equal vote to other stakeholders (Deverka *et al.*, 2019) and have a real influence on research. For example, if the group feels that some big data project does not have enough public input, they can assign a public contributor to support that particular work (Willison *et al.*, 2019). The governance bodies could also assist with engaging the general public (e.g. by reviewing lay information), and guide the recruitment of new public contributors (Jones *et al.*, 2020b). The influence of governance groups differs, and O'Doherty *et al.* (2011) recommend flexible governance that might evolve as big data research develops. Some papers argue that a one-size-fits-all solution might never work in big data research or for diverse communities (Bot, Wilbanks and Mangravite, 2019; Henare *et al.*, 2019; Mourby *et al.*, 2019; Tindana *et al.*, 2015). Embedding involvement in the governance of big data research may require novel solutions (Erikainen *et al.*, 2021).

Involvement around governing big data research could also be conducted as a one-off deliberation event (Teng *et al.*, 2019; Deverka *et al.*, 2019) or a Delphi study (Kirkham *et al.*, 2021). A one-off deliberation process could particularly benefit contentious issues (O'Doherty *et al.*, 2021). For example, Townson *et al.* (2020) organised workshops to co-produce a methodology for raising awareness of the research project that also aimed to improve recruitment.

5.3.2.2. Engagement

Engagement is about reaching the broader public, especially around dissemination (Mourby *et al.*, 2019; Kalkman *et al.*, 2019a). Engagement is mentioned alongside education as it is about showing how findings from big data projects are applied back to the community (Evans *et al.*, 2020). Educating the public can be seen as paternalistic, one-directional and top-down, so there is a need for two-way communication (Shaw, Sethi and Cassel, 2020; Willison *et al.*, 2019). Engagement goes beyond the dissemination of information and requires that data activities have been influenced in some way by public views (Xafis and Labude, 2019).

Engagement could demonstrate the public sector are a trustworthy steward of patient data (Ballantyne and Stewart, 2019). Sharing research findings should include reaching individuals with personalised research results; these need to be valuable and benefit individuals (e.g. they could go for health tests or make life changes that improve their health) (Dankar, Ptitsyn and Dankar, 2018). The public should receive understandable, educational and operable data on project outcomes (Omar Muhammad *et al.*, 2020). Engagement activities should be proportional to the nature and size of the project around big data research (Ballantyne and Stewart, 2019). Therefore, how these engagement activities looked highly differed between the included papers. The public could be reached through engagement events (Manrique de Lara and Peláez-Ballestas, 2020; Jones *et al.*, 2020b). Events were held with service users (Newburn *et al.*, 2020). Researchers attended and supported events, e.g. during the colorectal cancer awareness month (Beyer, Comstock and Seagren, 2010). Interactive (graphics, videos etc.) were used during exhibitions to raise public awareness (Sleigh and Vayena, 2021).

Researchers should share any discussion from governance groups with a broader public. (Gallier *et al.*, 2021; Shaw, Sethi and Cassel, 2020) These could be a brief online report of findings and key recommendations (Beyer, Comstock and Seagren, 2010). Using modern technology, researchers could create a registry or website where the public can see who has access to their data and for what purpose, or receive newsletters (Ballantyne and Style, 2017; Dankar, Ptitsyn and Dankar, 2018; Newburn *et al.*, 2020; O'Doherty *et al.*, 2011). Newburn *et al.* (2020) aimed to share their research on social media (Twitter and Facebook). Nation-wide campaigns could explain the benefits of big data research (Goytia *et al.*, 2018; Evans *et al.*, 2020; Sleigh and Vayena, 2021). This should be done in the language (e.g. indigenous) the public understands (Henare *et al.*, 2019). The public could be further reached through patient organisations (Omar Muhammad *et al.*, 2020; Newburn *et al.*, 2020), and researchers could share (yearly) updates jointly with them (Duchange *et al.*, 2014). Researchers should remain vigilant of lobbied topics and media discourse (Bharti *et al.*, 2021).

The public could be a part of the engagement process. Townson *et al.* (2020) mention the role of "Champions" who promoted the study in GPs, large public events (e.g. food festivals) reaching schools and support events organised by researchers. Another role they had was that of "supports". Similarly to champions, their role was to promote the research, but it took the form of a pledge; this was more casual, with no formal training or evaluation and no reimbursement. However, both roles were voluntary, with no specific targets to reach (Townson *et al.*, 2020).

5.3.2.3. Consultation

The consultation approach consists of a survey (Omar Muhammad *et al.*, 2020; Duchange *et al.*, 2014), informal small group meetings (e.g. town hall meetings) (Tindana *et al.*, 2015) or qualitative research aiming to capture the public perspective (Fernando, King and Sumathipala, 2019). These include focus groups (e.g. exploring patient's approach to patient engagement in governance and prioritising research questions), and interviews (e.g. to understand public views towards privacy) (Ohno-Machado *et al.*, 2014).

5.3.3. Outcomes

Included papers in the review claimed that involvement and engagement offered numerous outcomes. First, it could identify gaps in knowledge and priorities for research (Nelson and Burns, 2020). Second, it could align researchers' and institutional perspectives of public interest with public views (Bharti *et al.*, 2021), for example, by bringing together charity workers, service providers, elected politicians and members of the public (Nelson and Burns, 2020; Fleurence *et al.*, 2014). Third, public contributors involved in governing bodies could improve trust and accountability (Vayena and Blasimme, 2017). Fourth, improving public awareness of big data might democratise health research (Kalkman *et al.*, 2019a). Vayena and Blasimme (2017) argue further that a blending of citizen science and participatory models could offer more democracy in governance. Involvement was considered an answer to new data-sharing schemes in the UK rather than the opting-out consent model (Coulter, 2021).

However, measuring the impact of involvement and engagement in big data research is challenging (Luna Puerta *et al.*, 2020; Tindana *et al.*, 2015; Townson *et al.*, 2020; Newburn *et al.*, 2020). A scoping review by Luna Puerta *et al.* (2020) recognised that there is no consensus about the objectives of public involvement in big data, which undermines the ability to measure impact. Another review by Tindana *et al.* (2015) found that the papers included in the review on community engagement did not evaluate the effectiveness of their engagement activity.

5.3.4. References to involvement or engagement guidance

A minority of papers directly referred to involvement or engagement guidance**.** These included the UK National Standards for Public Involvement (Jewell *et al.*, 2019; Hurt *et al.*, 2019; Jones *et al.*, 2020b), NIHR definitions of involvement and engagement (Newburn *et al.*, 2020; Townson *et al.*, 2020), GRiPP2 checklist (Newburn *et al.*, 2020; Jewell *et al.*, 2019), the consensus statement on public involvement and engagement with data-Intensive health research (Jones *et al.*, 2020b), an academic model guiding involvement (Baart and Abma, 2011), and local policies or principles (Dankar, Ptitsyn and Dankar, 2018; Rowe *et al.*, 2021).

5.3.5. References to legal documents

Some papers mentioned legal documents to justify involvement and engagement. These include data protection legislation (Jones *et al.*, 2020b; Milne and Brayne, 2020), government policies (Ballantyne and Style, 2017; Bot, Wilbanks and Mangravite, 2019), legislation or treaties around indigenous communities' rights (Funnell *et al.*, 2020; Henare *et al.*, 2019).

5.3.6. Challenges to involvement and engagement

Deciding who should be on the advisory board, how they should be selected and what their role remains a challenge for researchers (Tindana *et al.*, 2015). Advisory groups might not be representative of the public views (Milne, Sorbie and Dixon-Woods, 2021; Murtagh *et al.*, 2018; Xafis and Labude, 2019). No single committee can represent all communities (because of their diversity) (Paprica *et al.*, 2020; Henare *et al.*, 2019). Identifying the relevant communities is difficult (Tindana *et al.*, 2015). This creates a challenge of ensuring legitimate group representation (O’Doherty *et al.*, 2011). Advisory groups do not reach a broader population (Mourby *et al.*, 2019), so involvement and engagement need to move away from usual suspects (Paprica *et al.*, 2020; Milne, Sorbie and Dixon-Woods, 2021; Jones *et al.*, 2020b; Jones, McNerney and Ford, 2014). There is the risk that more vocal individuals dominate the discussion (Tindana *et al.*, 2015). Public contributors could be chosen arbitrarily, for example, based on personal contracts, and thus the process might not be transparent to the public (O’Doherty *et al.*, 2011). This could lead to involving financially, politically motivated (Deverka *et al.*, 2019) or well-connected contributors (Ballantyne and Stewart, 2019). The way to overcome that is to recruit potential study participants as public contributors; this could include participants electing their own representatives or market company leading the recruitment (O’Doherty *et al.*, 2011; Teng *et al.*, 2019).

Meaningfully including public contributors in the networks’ governance could be challenging (Fleurence *et al.*, 2014). Big data is a complex topic to involve and engage public contributors with sufficient big data expertise (Evans *et al.*, 2020; Manrique de Lara and Peláez-Ballestas, 2020; Deverka *et al.*, 2019; Jones, McNerney and Ford, 2014; Baart and Abma, 2011; Goytia *et al.*, 2018; Dankar, Ptitsyn and Dankar, 2018). Potential contributors might feel apprehensive about contributing to complex research if they do not understand the technical jargon (Jones *et al.*, 2020b; Ballantyne and Stewart, 2019). This can be further compounded by language and cultural barriers between researchers and the public (Tindana *et al.*, 2015). Public contributors should be offered training and additional support as required, especially when the topic gets more complicated (Jewell *et al.*, 2019; Townson *et al.*, 2020). Support needs to be person-centred and based on each individual skills and experience (Townson *et al.*, 2020). These could include short lectures, group discussions and opportunities to ask questions

(Jewell *et al.*, 2019; Milne, Sorbie and Dixon-Woods, 2021). For example, Teng *et al.* (2019) sent a booklet written by researchers in lay language on big data with a special focus on data collection, regulation, data sharing and public concerns. Involving people with experience in research could be an alternative (Murtagh *et al.*, 2018). Kirkham *et al.* (2021) included public contributors with big data research experience. Still, the authors recognise that people with more understanding of big data might have different views than the general public.

Conducting involvement and engagement activities does not mean that public values are incorporated into big data research (O'Doherty *et al.*, 2011). Involvement could be tokenistic without real change, but this still could offer some form of legitimacy to researchers (O'Doherty *et al.*, 2011). There is a need to ensure power-relationship between public contributors and the research team (de Freitas *et al.*, 2021). When public contributors join already ongoing research projects, they have limited scope for impact (e.g. only ethics amendments might be allowed); thus, their involvement might turn more into consultation (Tindana *et al.*, 2015; Milne, Sorbie and Dixon-Woods, 2021). Some researchers do not support involvement and use a deficit engagement model (Baart and Abma, 2011). Researchers should reflect on how to ensure balance in engagement. It is about raising awareness of big data research and understanding the public views but should not be limited to an already agreed outcome but rather an ongoing dialogue (Jones *et al.*, 2020b; Jones *et al.*, 2019; Paprica *et al.*, 2020). Public involvement and engagement should take place before any data sharing occurs (Shaw, Sethi and Cassel, 2020).

Further barriers to PPI are around the organisational level. Secrecy could be a challenge (Shaw, Sethi and Cassel, 2020). Organisations might not want to share controversial information, and private companies may argue that sharing it might be again their commercial interest (Ballantyne and Stewart, 2019). Involvement and engagement have the potential to improve public trust in big data research but not necessarily in the research institution (Erikainen *et al.*, 2021). There is historic mistrust from underserved communities, e.g. African Americans, indigenous communities and people living with homelessness (Shaw, Sethi and Cassel, 2020). There is no guarantee that it is always possible to maintain public trust in big data research (Milne and Brayne, 2020).

There are practical barriers. In-person activities can be time and cost-restrictive for some communities (Ohno-Machado *et al.*, 2014). Public contributors might not attend meetings without warning due to personal circumstances (e.g. health treatment, work or family responsibilities) (Teng *et al.*, 2019; Newburn *et al.*, 2020).

5.3.7. Facilitators of involvement and engagement

Various ways could be utilised to reach diverse audiences (Omar Muhammad *et al.*, 2020; Willison *et al.*, 2019). Recruitment of public contributors is mostly through already existing groups like involvement groups or patients organisations (Jones, McNerney and Ford, 2014; Weich *et al.*, 2018; Omar Muhammad *et al.*, 2020; Jewell *et al.*, 2019; Ohno-Machado *et al.*, 2014), clinical sites (Ohno-Machado *et al.*, 2014), or recruitment via newsletter distributed among study participants (Townson *et al.*, 2020; Hurt *et al.*, 2019). Working with intermediaries (e.g. charities or community leaders) could improve the success of reach as they provide advice around public perspective or become gatekeepers (Nelson and Burns, 2020; Tindana *et al.*, 2015). Public contributors might be unclear on their role at the beginning (Jones, McNerney and Ford, 2014). Therefore, clear criteria for public contributors are needed (Milne, Sorbie and Dixon-Woods, 2021). Selling involvement should focus on being a reciprocal opportunity with benefits for both researchers and public contributors (Newburn *et al.*, 2020). The recruitment ad should include the role description and required skills (Jewell *et al.*, 2019). The full research protocol might be available on request (Newburn *et al.*, 2020). There is a need for a transparent process of selecting public contributors to avoid tokenism (Deverka *et al.*, 2019; O'Doherty *et al.*, 2021). Candidates could be interviewed to identify individuals with team working skills and the ability to contribute outside their own health situation (Willison *et al.*, 2019; Jones, McNerney and Ford, 2014) as public contributors' emotional connection to the research could be both an enabler or a barrier to their involvement (Nunn *et al.*, 2021).

Researchers should communicate clearly, in lay language and without jargon to ensure transparency (Paprica *et al.*, 2020; Deverka *et al.*, 2019; Jones, McNerney and Ford, 2014). The examples include jargon-free graphics (Sleigh and Vayena, 2021), tailoring academic research to lay audience (Baart and Abma, 2011), and post-session informal debrief (Murtagh *et al.*, 2018). When reaching the broader public, researchers should aim to deliver the message themselves rather than through the lens of media to provide more balanced information (Newburn *et al.*, 2020). Public contributors receive training introducing them to big data research (Willison *et al.*, 2019; Jones, McNerney and Ford, 2014; Townson *et al.*, 2020; Murtagh *et al.*, 2018; de Freitas *et al.*, 2021). The availability of good-quality information on big data underpins meaningful public involvement (Xafis and Labude, 2019; Omar Muhammad *et al.*, 2020). Explanations could include links to Wikipedia (Townson *et al.*, 2020). Researchers should send information prior to activities to give people time to reflect on it (Townson *et al.*, 2020). Public contributors might need extra time to consider their responses (Jones *et al.*, 2020b). Before meeting other stakeholders, public contributors can meet first together (de Freitas *et al.*, 2021). When commenting on a new aspect of research, public contributors are invited to comment first (Willison *et al.*, 2019). Some papers described the beginning of the involvement process (Baart and Abma, 2011; Teng *et al.*, 2019). In Teng *et al.* (2019), during the first day of activities, presentations were made to provide background to public contributors. These were from the perspective of a patient and seldom-heard community. These presentations were not neutral but opinionated to show the diversity of views on big data research.

Expectations around monetary compensation should be set up from the start (Tindana *et al.*, 2015). These include reimbursement for time (Jewell *et al.*, 2019; O'Doherty *et al.*, 2011; Townson *et al.*, 2020; Teng *et al.*, 2019), travel (Teng *et al.*, 2019) and childcare expenses (Newburn *et al.*, 2020). Researchers should provide lunch (Newburn *et al.*, 2020) and use venues easily accessible by public transport (Newburn *et al.*, 2020). If public contributors are paid equally to professionals in governing bodies, this might improve their involvement (Deverka *et al.*, 2019).

Involvement and engagement are led by team members experienced in or organising and running these activities (Nelson and Burns, 2020; Hurt *et al.*, 2019; Jones *et al.*, 2020b; Paprica *et al.*, 2020; de Freitas *et al.*, 2021). Other researchers should dedicate time to these activities (and this time should be embedded in the workload) (Jones *et al.*, 2020b). Research team members and facilitators are trained in public involvement (Hurt *et al.*, 2019; Teng *et al.*, 2019). Access to specialist training on involvement and engagement should be provided to both staff and the public (Jones *et al.*, 2020b).

Researchers should respect local and seldom-heard groups' traditional structures and ethical perspectives. Papers focusing on indigenous communities showed already existing governance mechanisms supporting research with these groups (Rowe *et al.*, 2021; Hudson *et al.*, 2020). Researchers should incorporate indigenous culture, e.g. traditional ceremonies, when reaching the community (Henare *et al.*, 2019). Formalised agreements with indigenous organisations could improve the relationship with that community (Funnell *et al.*, 2020). This more nuanced approach to big data research could assist researchers in establishing trust with indigenous communities rather than merely convincing them this is the right thing to do (Hudson *et al.*, 2020).

Respectful, ongoing, genuine, and non-hierarchical interaction between researchers and the public is needed to build trust (Xafis and Labude, 2019; Jones *et al.*, 2020b). Building a relationship takes time (Tindana *et al.*, 2015). It includes the co-ownership of research (Funnell *et al.*, 2020) and concentrates on what the public wants to know (Baart and Abma, 2011). The reciprocal relationship is illustrated by Newburn *et al.* (2020), who organised workshops during which they delivered training for members of the public on using social media, policy updates and research methodology. A clear purpose for activity leads to realistic expectations (Jones *et al.*, 2020b). The starting point of involvement might not be about an equal partnership but an exploration of power relationships (Baart and Abma, 2011). Working in smaller groups gives more opportunities to every public contributor to share their opinion (Teng *et al.*, 2019). Decisions could be made through consensus (Funnell *et al.*, 2020; Willison *et al.*, 2019). However, Ballantyne and Stewart (2019) recognise that there will always be disagreements; in that case, there is a need for a clear explanation for why these voices were not included.

5.4. Stage 2: Systematic Review

Out of the papers included in the scoping review, four papers were identified discussing the delivery or effectiveness of involvement and engagement strategies (Jewell *et al.*, 2019; Sleigh and Vayena, 2021; Jones, McNerney and Ford, 2014; Jones *et al.*, 2020b). The rest were excluded because they did not discuss the delivery or effectiveness of involvement or engagement strategies (n=45) or were not empirical studies (n=4). The majority of papers included in the scoping review (stage 1) were theoretical or a description of what they expected from the involvement or engagement activities rather than reporting what took place.

Out of included papers in the systematic review, three of them are small evaluations of ongoing involvement and engagement activities (Jewell *et al.*, 2019; Jones, McNerney and Ford, 2014; Jones *et al.*, 2020b) and one descriptive case study (Sleigh and Vayena, 2021). The evaluation methods were group discussion (n=1) or survey (n=3). All of them were done by researchers directly involved with public contributors. As two papers present the case of SAIL Databank (Jones, McNerney and Ford, 2014; Jones *et al.*, 2020b), the four papers cover only three distinctive initiatives.

As only four papers were identified, two of them examined the same involvement activities, and only small sections of included papers focused on research on PPIE, it was impossible to conduct qualitative synthesis. As this systematic review is empty (not enough data is available to conduct a systematic review), it indicates the need for more research that examines the delivery and effectiveness of involvement and engagement activities.

5.5. Discussion

The scoping review provides an overview of evidence around public involvement and engagement in big data research. This is the first review exploring this issue. The review has shown that the public can and should be involved and engaged in big data research in terms of individual initiatives and data governance. However, the findings indicate that there is no one right way to involve and engage the public in big data research. Researchers should consider what kind of activities apply to their work and should use multiple approaches (involvement, engagement and consultations) to reach different communities. Some papers suggest utilising modern technology when engaging the public (e.g. through a website or newsletter). Surprisingly, there has been limited discussion around social media as a potential avenue to reach the broader public.

The review indicates that public involvement and engagement have the potential to improve public trust and accountability for data resharing for research. However, there is no literature actually evaluating these activities. Future research should attempt to measure the impact of involvement and engagement in securing the social licence for big data research with the broader public.

Papers mentioned seldom-heard community and vulnerable groups and argued that involving and engaging them in big data research is important. However, there is limited evidence of how this should look in practice. First, no included paper conducted any study exploring the process of involvement and engagement from the perspective of researchers. Second, the papers that examined public contributors' perspectives included mostly people who were already involved as public contributors.

References to PPIE guidance or legal documents in the included papers were limited. The consensus statement on public involvement and engagement with data-intensive health research (Aitken *et al.*, 2019) is relatively new guidance. However, INVOLVE has been active in the UK since 1996. This indicates that many included papers replicate similar discussions around principles involving and engaging the public rather than referring to already established standards.

The findings indicate that some challenges are particular to involvement and engagement around big data research. The main one is that big data research remains a complex topic. It might not be easy to briefly (or in accessible language) explain it to potential public contributors or the general public. However, the papers offered some suggestions on how these barriers could be overcome. Researchers need to build time and resources when discussing big data research with members of the public. However, this experience overlaps with another review that looks at patient involvement in cancer research, as the authors found that the main challenge of involvement was that it was time-consuming (Pii *et al.*, 2019). My review suggests that involvement and engagement of the public in big data research might be even more time-consuming. The other specific issue around involvement in big data research were selecting public contributors to sit on governance bodies. If these challenges are overcome, there is a higher chance that involvement and engagement in big data research is not tokenistic.

5.6. Revised model

The previous chapter presented the system model for public involvement and engagement in big data research.[17] This has been used in two ways. First, based on the review findings, the model was developed. Figure 4 presents the updated version. Second, using the model, I identified issues that were not yet discussed in the literature and reflected if the model could apply to public involvement and engagement in big data research.

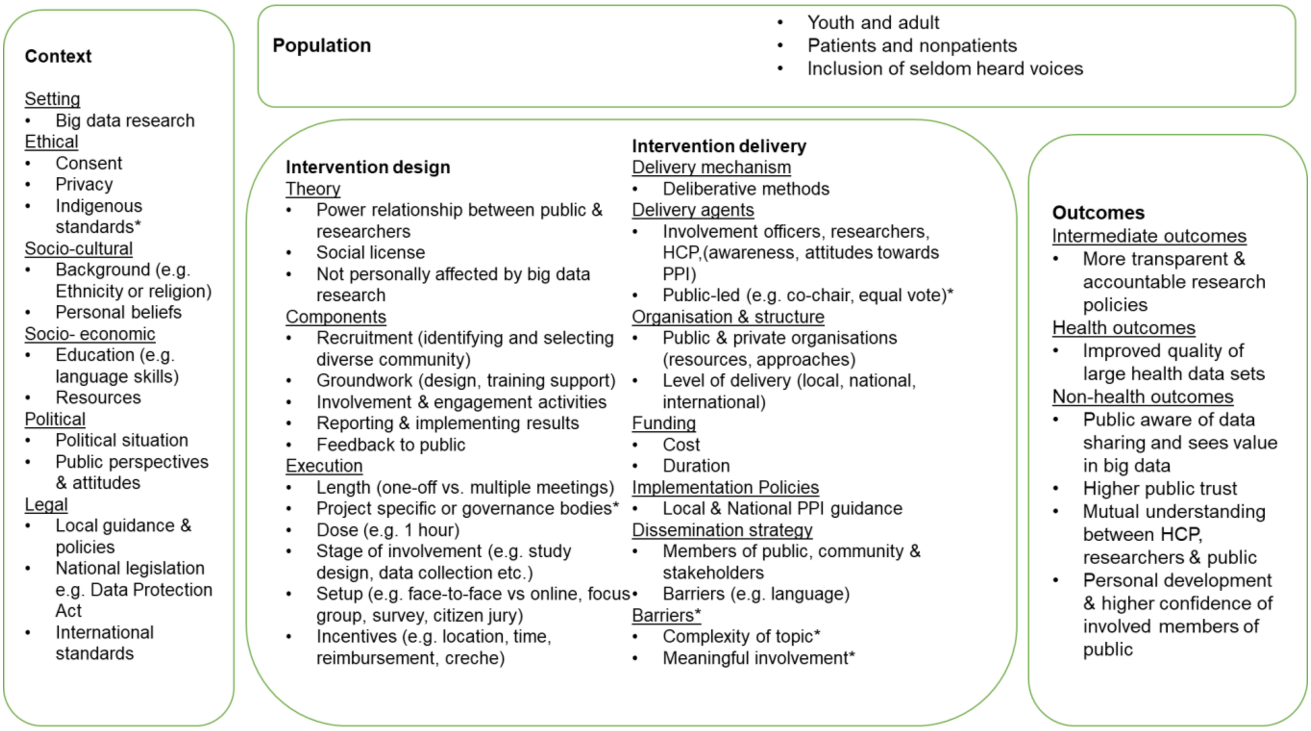


*Figure 4 The updated system logic model of public involvement and engagement in big data research. Asterix * is used to record new aspects of the model based on the review.*

Within the context section, indigenous standards were added to recognise that big data research needs to consider the perspective and views of indigenous communities that might differ from white Western perspectives. In the intervention theory, I included that execution of involvement activities could be divided into project-specific (e.g. focusing on one big data research project) or governance bodies that look into granting approvals into data linkage (for other projects). These two purposes might influence how researchers involve and engage the public. In intervention delivery, I added the bullet point around public-led activities. Some papers suggest ensuring that the public voice is equivalent to professionals during voting and has more influence (e.g. by co-chairing meetings or being co-investigators). Furthermore, a new bullet point was added in intervention delivery to recognise big data-specific barriers, especially jargon, and how complex big data research could be to members of the public.

---

[17] See the initial system logic model at page 63.

The majority of the aspects of the model were discussed in the included papers. The only exception is not reflecting on the involvement and engagement of people who are not personally affected by big data research (or do not perceive themselves as such). This might be explained by a limited number of papers exploring the perspectives of members of the public (as they were discussions rather than primary research). Future research should explore more public perspectives, especially through qualitative research. The coverage of the majority of issues for the involvement and engagement in big data research suggests that model could support researchers who intend to design and deliver these activities to the public.

5.7. Limitations

The review explored public involvement and engagement in big data research. These terms are used differently by researchers.[18] This could have influenced the search results as potentially some relevant papers might not have been picked up by the search strategy as the authors would use different terms. I developed the search strategy with an experienced librarian and conducted additional manual searches. However, this did not guarantee that all relevant papers were included. This is a similar experience to Brett *et al.* (2014) review, as they felt that the variability of wording used to describe involvement led them to a complex search. The second limitation was that only information included in the papers was extracted. I did not approach the authors for more detail. As academic papers have a word limit, it is possible that some additional information about involvement and engagement may have been missed.

5.8. Conclusion

This chapter presented the findings from the scoping and systematic reviews. First, I discussed methods and reflected on how these followed the published protocol. Then, I reported the findings from this two-stage review. Later, I discussed the importance of these findings and developed the model for public involvement and engagement in big data research.

The review indicates areas for future research. Three of them will be explored in this thesis. First, the majority of included papers were discussion papers, and there is a need for more research to explore the views of researchers who involve and engage the public in big data research. Second, there is little known from the perspective of members of the public on how to involve and engage them. This is especially the case from the perspective of seldom-heard communities and people who have not been previously involved in big data research. Third, some papers encouraged the use of modern technology to engage the public, but there

---

[18] See more in chapter 1, at page 23 for involvement and big data at page 27.

remains limited evidence on how to do it (especially on social media). These areas of future research underpin the reasons for conducting the three qualitative studies presented in this thesis. Hence, I explored how Twitter was used to engage the public and conducted interviews with researchers and seldom-heard communities. The following chapters discuss the methodology, data analysis and public involvement in my research, and then I report findings from studies to address these gaps in the literature.

6. Methodology

6.1. Introduction

The overarching aim of this chapter is to discuss the philosophical stance, methodology and methods used in this thesis that will not be included in the later chapters reporting research findings (as these were written for peer-reviewed academic publications). First, I present the philosophical stance that informed the methodological choices made in this thesis. I define and justify my decision to use qualitative methodologies and methods to answer the research questions. This will be linked with the discussion on sampling and recruitment of participants. Then, I provide a step-by-step account of the process for conducting interviews, and later I reflect on my positionality as a researcher. Finally, I discuss the ethical issues raised when conducting this research.

6.2. Philosophical Underpinnings

This thesis takes social constructionism stance, looking at how participants have constructed PPIE. Therefore, the section introduces the social constructionism approach to social science research and then discusses its features. As the social constructionism epistemological and ontological stance informs my decisions about research methods and approaches to data analysis, I explain where I drew on this philosophical stance in my research in subsequent sections and chapters where they are further developed.

6.2.1. Social constructionism as an approach to social science research

Social constructionism has been used since the 1960s in social research (Gubrium and Holstein, 2008). Its prominence started with Peter Berger and Thomas Luckmann's 1966 book, the Social Construction of Reality. They explored how subjective meaning became a social fact and proposed the theory that people and how they interact create society. For example, we see the hospital as a place where people with sicknesses receive medical assistance, not because of a building but because there has been a consensus among people (often people coming before us) that this building acts as a hospital.

Others have further developed Berger's and Luckmann's ideas. Social constructionism remains a broad movement (Galbin, 2014; Braun and Clarke, 2021c). Its development has been influenced by different philosophies, such as feminism and poststructuralism (to name a few) (Burr, 2015). I especially draw on works by Kenneth Gergen (1973; 1989; 2015). In his essay on "Social Psychology as History" (Gergen, 1973), he challenged the assumption that we can find final descriptions of the world. He claimed that knowledge is culturally and historically constructed depending on time and context. What it means for researchers adopting this stance is that according to the social constructionism world view, all meaningful reality is socially constructed (Gergen, 2015). The world is not just out there but rather constructed by people (Gubrium and Holstein, 2008). Culture, history and context can influence how knowledge is constructed. Bryman (2016, p. 29) adds that

> "*it implies that social phenomena are not only produced through social interaction but are in a constant state of revision*."

In other words, since we were born, our interactions with people, the community and the environment around us have influenced how we construct (and thus perceive) our reality (Galbin, 2014). The new experiences keep constructing (and refining) our perspectives (Schwandt, 2000).

Social constructionism shapes both epistemological and ontological perspectives. Edwards (1997, pp. 47-48) describes that as

> "[a] *mind is 'socially constructed' ontologically (…) mind is real for the theorist and analyst, and the analytic task is to explain how it is built within a real world of cultural settings and practices. (…) 'social construction' is epistemic; it is about the constructive nature of descriptions, rather than of the entities that (according to descriptions) exist beyond them*".

Edley (2001, p. 437) further builds on that stance

> *"as soon we begin to think or talk about the world, we also necessarily begin to represent (…) Talk involved the creating or construction of particular accounts or stories of what the world is like."*

Burr (2015, p. 105) builds on both Edwards' and Edley's works by proving the clear meaning of the ontological stance:

> "*When used ontologically the term social constructionism refers to the way that real phenomena, our perceptions and experiences, are brought into existence and take the particular form that they do because of the language that we share. This does not make these phenomena or things unreal, fictitious or illusory: they are no less real for being the products of social construction*."

These quotes show that social constructionism takes a critical stance toward taken-for-granted knowledge. It undermines positivists' confidence in the truth and reality (Edley, 2001). The next subsection will discuss what it means in practice for researchers, as I introduce key features of the social constructionism way of thinking.

6.2.2. Features of social constructionism

What constitutes social constructionism (sometimes even in terms of epistemological and ontological positions) differs among researchers (Edley, 2001). Gubrium and Holstein (2008, p. 5) call it a "*mosaic of research*", thus emphasising this diversity of views, but they also recognise that this mosaic shares common features. The term social constructionism is even often interchangeably used with constructivism (e.g. in work by Charmaz (2000; 2006)). Others prefer the separation of these terms (Pfadenhauer, 2018). I follow key features as identified by Burr's and Gergen's works. Burr (2015, p. 2), following Gergen's research, uses the term "*family resemblance*" to identify key assumptions that social constructionism approaches could share. I discuss each of these assumptions recognising that social constructionism has emerged as a response to positivism and empiricism.

1. A critical stance toward taken-for-granted knowledge

Social constructionism challenges positivistic thinking that our understanding can be based on objective and unbiased observation of the phenomenon. However, this does not mean a complete rejection of the rational world but rather recognises that our perspectives are subjective (relative) rather than absolute (Kamberelis and Dimitriadis, 2016). The argument is based on the assumption that one group of subjective humans (researchers) studies the other group of subjective humans (participants) (Galbin, 2014). Researchers are part of the culture and social norms, so their interpretations of data will be guided by their experiences. They live in the community and have political views and clear moral standpoints. Claiming that research can be morally and politically unbiased can be misleading (Gergen, 2015). To overcome this challenge in the thesis, I shall acknowledge my own biases and critically reflect on them in the positionality section.[19]

2. Historical and cultural specificity

Positivism looks for universal truth, claiming that by understanding participants' experiences, we can apply them to the wider population. However, social constructionism claims that individuals' understanding of the world is historically and culturally specific. Living in different societies could bring dissimilar consequences; in theory, concepts like money or citizenship could have been constructed differently (Boghossian, 2001). As presented in the brief history of PPI in healthcare services and research[20], the meaning of PPI has changed over time and still heavily differs among research teams, disciplines and research funders. Also, what involvement is (or if it is practised at all) could vary among countries. People's actions are

---

[19] See section on reflexivity and positionality at page 116.

[20] See the history of public involvement in the UK at page 30.

constantly in motion, so predicting them remains challenging (if not impossible). In my research, I do not seek a universal truth but rather understand how participants as individuals express or shape their experiences, views and perspectives towards public involvement and engagement. As the results of this thesis are socially constructed and based on participants' worldviews, they cannot be generalised to all Polish or South Asians communities in the UK, researchers in big data research, or Twitter users.

Implementing change in a constantly (and rapidly) transforming world can be challenging. Social constructionism can offer practical insights into the policy (Conrad and Barker, 2010). Changing people's perspective (their construction) of the phenomenon can change how others perceive it. For example, this can be seen in medical sociology. If one perceives obesity as a risk factor (for sicknesses like diabetes), then policy responses are limited to exploring the role of food or access to healthy food choices. However, if obesity is an illness, patients could be offered treatment (such as gastric bypass surgeries) (Conrad and Barker, 2010).

3. Knowledge sustained by social processes

People construct their worldviews among themselves through their daily interactions with others. Referring to the evolution of PPI in healthcare[21], some researchers have now recognised PPI as an essential part of health research. However, some researchers remain sceptical or adamant towards involving patients in research as they are not trained researchers (Boylan *et al.*, 2019). Different levels of involvement can be identified in Arnstein's ladders of participation. Thus, the existence (and role) of public involvement in research remains fluid; it is a social process constantly influenced by people's interactions, culture and history. The assumptions of social constructionism are not limited to data or participants' realities but also to the researcher's version of reality (Bryman, 2016). My understanding and presentation of data will be one of many possible realities, and I do not see it as definitive. In other words, social constructionism does not aim to persuade others about how the phenomenon should be understood but to offer multiple ways of looking at it (Galbin, 2014).

4. Knowledge and social action as going together

Social constructionism encourages us to deconstruct traditional practices and reflect on them (Galbin, 2014). New constructions of the world can bring different (positive) social actions. For example, in the past, researchers perceived people living with disabilities as unable to participate fully in society. Nowadays, it has changed, and social barriers, not their impairment, are seen as impacting people living with disabilities. The construction of why people with

---

[21] See literature review section at page NUMBER

disabilities socialise less or are less often employed has not changed, but our construction of the phenomenon has. This change would not have been possible without activists who shaped that perspective. Social constructionism encourages collaborative inquiry (Gergen, 2015) to pursue change. It encourages a move from a researcher-only approach to more co-production of knowledge, leading to real-life implications (Galbin, 2014). Researchers and public contributors can jointly construct the reality that often can bring new social actions. In my thesis, two public contributors were involved and supported the research process. The chapter on public involvement within my research discuss this in detail.[22]

This section has introduced social constructionism and briefly highlighted how it was used in this thesis. As Gergen (2015, p. 61) points out, social constructionism "*invites us to appreciate multiple perspectives*". PPIE is not a fixed series of stages; each public contributor, member of the public or researcher could perceive it differently. Thus, there is no one truth, multiple perspectives of PPIE can exist simultaneously, and each is equally viable. Meaningful PPIE will require exploring this diversity of views. Correspondingly, this thesis adopted social constructionism to unpick these realities and understand how the public can be involved and engaged in big data research. Social constructionism enables the exploration of subjective reality. Qualitative methodology is often suitable for studying the subjective experience of individuals. Social constructionism has been successfully used in qualitative research (Gubrium and Koro-Ljungberg, 2005; Braun, 2008). Therefore, in the next section, I explore qualitative research.

6.3. Qualitative research

In the previous section, I discussed social constructionism; as the appropriate methodology for that philosophical stance is qualitative inquiry, this section discusses the key characteristics of qualitative research. This is achieved by comparing qualitative with quantitative research. Although contrasting qualitative with quantitative research has its drawbacks, qualitative work became assessed in terms of what quantitative work is not (Bryman, 2016). My intention here is not to show which methodology is better but to emphasise why qualitative research was more beneficial in answering the research questions addressed in this thesis.

Qualitative research is a broad field that draws on diverse designs (Creswell, 2017). It would not be feasible to discuss all of them. Thus, I outline five key preoccupations of qualitative research as identified by Bryman (2016).

---

[22] See chapter 8 at page 143.

1) seeing through the eyes of the people being studied

Researching objects (e.g. natural sciences) and people differs as stone, chemicals etc., cannot tell us anything about their experiences, but people can. Thus, a qualitative researcher would argue that we need to understand the world from the perspective of people (who become study participants) (Bryman, 2016). This often requires exploring participants in their natural settings through extended face-to-face interaction over time (Creswell, 2017). Quantitative research is more often in controlled settings where researchers would fully control external factors. Quantitative researchers are distanced from their participants. However, being in natural settings impacts qualitative researchers, who could become influenced by their personal experiences, opinions or feelings in how they analyse and report the research findings. Reflexivity requires qualitative researchers to reflect on their backgrounds, views, experiences and beliefs and how they shaped their interpretations of data.[23] This is especially important as the researcher is a key instrument in collecting and analysing data, even if they are supported by programmes (e.g. NVivo)[24].

2) description and the emphasis on context

Qualitative research often presents a detailed description of the study area, such as participants or location, which some might consider "trivial details" (Bryman, 2016, p. 394). These descriptions allow the reader to understand the context of the participants' experiences. In other words, exploring participants' social world without understanding it first is impossible.

3) focus on the process

Qualitative researchers tend to perceive the social world as a process. They show how events develop and change over time. This requires a holistic account of the phenomenon (Creswell, 2017) and unfolding the events is not always presented as a linear progression but recognises multiple causes and effects between what participants do and their social world. The process can be studied in real-time or retrospectively (Bryman, 2016). The former approach is about collecting data as the events unfold (e.g. ethnographic research). The latter refers to when participants refer to past events (e.g. during interviews).

4) flexibility and limited structure

Qualitative research offers flexibility in how researchers perceive the world so they can understand it better from the participants' perspective rather than imposing any framework (through researchers' perspective, theory or previous research). This is linked with the

---

[23] See the reflection on my positionality on page 117.

[24] See my discussion around use of Nvivo in analysis at page 129.

principle of limited structure. Qualitative researchers often prefer to keep their research methods open to accommodate new and unexpected avenues of inquiry. For example, semi-structured or unstructured interviews give participants space to discuss things related to the research topic that researchers did not previously consider when designing a topic guide. Data derived from qualitative research is rich in detail and often presented in the form of words (but could also include videos, drawings or pictures). In contrast, quantitative research favours numbers. Both ways of presenting data have their merits. Social constructionism supports using statistics or numbers but sees them as another way of presenting the world around us. Thus, neither words nor numbers are a more precise way of perceiving reality (Gergen, 2015).

5) theory grounded in data

Concepts and theories in qualitative research are usually derived inductively from data. Grounded theory (Charmaz, 2006; Strauss and Corbin, 1997) would be a typical example of that process when a researcher follows systematic inductive methods to develop a theory based on the analysed data. Inductive thematic analysis (Braun and Clarke, 2021c) could also be conducted where the data would allow for determining the themes. Although, as the authors point out, themes do not simply ‘emerge’ (ibid). The role of a qualitative researcher is not to bring to the surface results which are hidden (like an archaeologist would discover new artefacts). This is rather an active process of analysis where results are produced through a mix of the researcher’s experience, disciplinary perspective, theories, concepts and the data itself. Quantitative research is based on deductive reasoning, where a researcher tests theories and concepts in the data. Qualitative research could be deductive too, but the inductive approach allows more flexibility to make more bottom-up meaning from the data. However, inductive and deductive approaches do not exclude each other in qualitative research. In fact, Creswell (2017) argues for always starting the qualitative analysis inductively and then moving to a deductive approach to see if there is enough evidence to support each finding.

In this section, I have shown the strength of qualitative inquiry. This thesis aimed to explore how seldom-heard communities can be involved and engaged in big data research. Qualitative research offered an opportunity to provide rich descriptions of participants’ perspectives on how the process looks and how seldom-heard communities can be reached more meaningfully. Thus, qualitative research was the most appropriate approach to explore this issue. As I have chosen the qualitative inquiry for this thesis, the next issue to consider was which methods would be most appropriate to answer the research questions. This is discussed in the next section.

6.3.1. Methods

As discussed in the previous section, the qualitative methodology consists of multiple research methods. This section outlines the research methods chosen to empirically investigate how seldom-heard communities can be involved and engaged in big data research. These methods have been selected for their ability to answer the thesis research questions from social constructionism perspective. Involving and engaging public members (including seldom-heard communities) in big data research is complex. Therefore, this thesis looked at the phenomenon from multiple perspectives. These included qualitative interviews with Polish and South Asian participants (as subgroups of seldom-heard communities), researchers (including PPIE facilitators), and data collected from Twitter. This approach was appropriate for social constructionism, which requires exploring different perspectives and understanding how various actors construct the phenomenon (Gergen, 2015); in the case of this thesis, involvement and engagement.

First, I explore qualitative interviews with a special focus on conducting them in different languages and online. Thereafter, I discuss my sampling and recruitment strategies. I complete the discussion around qualitative interviews by providing a step-by-step description of how the interviews took place. Then, I examine conducting research on Twitter. Finally, I discuss data saturation for both the interviews and Twitter data.

6.3.1.1. Qualitative interviewing

Qualitative interviews, also known as in-depth or intensive interviews, provide a richness of data through an in-depth exploration of interviewees' experiences through participants' lenses and often offer considerable insight into the studied phenomenon (Charmaz and Belgrave, 2012). At least two people have to participate in it: an interviewer (a researcher) and an interviewee (a participant). Interviews are one of the most popular methods in qualitative research, and Rubin and Rubin (2005, p. 12) call them "*extensions of ordinary conversations*." In practice, interviews can also be unpredictable, with participants trying to reassess some form of control, for instance, by changing the subject of conversation. This is the role of researcher to ensure that discussion remains on the relevant topic.

There are three types of qualitative interviews based on the structure of the interview guide. In a structured interview, a researcher uses the topic guide with the same fixed number of questions to each participant, following the same order. A semi-structured interview offers more flexibility as there is a topic guide to support the researcher, ensuring that each participant covers the same discussion area. However, there is no fixed number of questions. The topic guide can include prompts, and the researcher can change the order of questions or add a new follow-up if the participant starts to explore a new avenue around the research aim. Last, unstructured interviews would start with one question at the beginning of the conversation and flow much more naturally than the other two types. Here, the interviewee has more freedom to elaborate on issues of their choice.

Semi-structured interviews were deemed the most appropriate to address this thesis's research questions. There has been limited literature on the involvement and engagement of members of the public in big data research; hence it would be difficult to design a structured interview. Thus, more flexibility during the interview was required to provide a richness of data. On the other hand, unstructured interview questions are very time-consuming for potential participants. Due to their length, unstructured interviews tend to be very long, which could lead to a small sample size. Semi-structured interviews offered an opportunity to overcome these challenges.

6.3.1.1.1. Interviews in different languages

Some participants in my thesis included people for whom English is a second language.[25] Therefore, some data collection was conducted in their mother tongue. Conducting research in one language and reporting in another can influence its validity, and differences can appear in meanings during translation (Gawlewicz, 2016). Words or concepts used in one language might not have an equivalent in the other (Gonzalez and Lincoln, 2006). Therefore, qualitative researchers need to understand the challenges of conducting research in different languages and consider strategies to overcome them. I use two terms: interpretation and translation. In this thesis, they have different meaning; despite the fact that some literature used them interchangeably (Temple, 2002). Interpretation deals with spoken language, while translation focuses on written content. In this subsection, first, I explore the experience of interviewing participants in one language and reporting the findings in another. Second, I briefly discuss conducting qualitative interviews with the support of interpreters.

English is the language of academia (e.g. dissemination of research is mostly done in English) (Kara, 2018), but many researchers conduct their work in other languages. This discrepancy remains a concern for qualitative researchers (Gawlewicz, 2016). As a native Polish-speaking researcher, I offered Polish participants a choice to have the interview in their mother tongue. This meant that data had to be translated for inclusion in the thesis and in any other publication. During the translation process from Polish to English, I was concerned that the essence of the message shared by a participant would be unintentionally lost, as can be the case when data is not reported in the participants' language (Squires, 2008). In other words, what would be clear in the original, might have a different or limited meaning in English. To

---

[25] See more details about participants please see the section on sampling at page 105.

overcome this challenge, I followed the guidance developed by Gawlewicz (2016) to note any potential discrepancies. Gawlewicz’s (2016) work is of special relevance to me as her research also focused on interviews in Polish. Thus, we shared the same language issues around Polish and English grammar differences, which can be challenging, if not impossible, to translate. Furthermore, we both are not qualified in linguistics. The process included creating notes when there were unsaid messages, the response was culturally influenced, or hard-to-translate words were used. The following extract presents this process:

> "*Well, we are in a foreign country, we are not in Poland, we are not citizens ... I mean a citizen. We are immigrants. [laughing] We are not British citizens, we have these statuses [such as settled status; PL: mamy te statusy]. Yes, we have the same rights as the British, but really, but despite everything I do not know how and I know many people too, always have the back of my mind [as back of my head] that we are not British and that it is not quite like that [to the end; PL: nie jest tak do końca], it is our home, but not our country*." (Marlena, translated from Polish)

> „*No jesteśmy na obczyźnie, tak nie jesteśmy w Polsce nie jesteśmy obywatelami... tak znaczy obywatelem. Jesteśmy imigrantami. [laughing] Nie jesteśmy obywatelami Wielkiej Brytanii, mamy te statusy. Tak mamy te same prawa co co Brytyjczycy to tak naprawdę, ale mimo Wszystko nie wiem, jak wiem, że u wielu osób u mnie zresztą też zawsze z tyłu głowy jest to, że no nie jesteśmy Brytyjczykami i że to nie jest tak do końca, to jest nasz dom, ale nie nasz kraj*.” (Marlena, Polish Original)

In contrast to Gawlewicz (2016), I followed this procedure only for quotes used in this thesis and publications as I analysed the interviews myself. As I have been living in the UK for over ten years and mostly speak and write in English, my Polish skills might have deteriorated. This is often expected of people who do not practice the language (Leinonen and Tandefelt, 2007). This translation process helped me reflect on how to stay close to the interviewees’ perspectives of the world when presenting data to English-speaking audiences.

The presence of interpreters offers an opportunity for easier communication with participants who might not feel comfortable expressing themselves in English (and where the researcher does not speak the interviewees’ language). Translators (and, by extension, interpreters) have a profound influence on shaping data (Gawlewicz, 2016) and, when not used correctly, can lead to poorer data quality if they do not follow research standards or cause miscommunication (Kosny *et al.*, 2014). Thus, the researcher should carefully plan their involvement in research (Squires, 2008).

Interpreters have different styles, as there are three types of interpreting: simultaneous, consecutive and whispering. In simultaneous interpreting, both the speaker and the interpreter speak at the same time. In consecutive interpreting, only one person speaks at the time, and often interpreter takes notes to ensure the quality of interpreting. Whispering is similar to simultaneous interpreting, but usually, the interpreter whispers the answer to one person.

An interpreter's experience can influence interpreting quality and even negatively impact the data quality (Rosenberg, Seller and Leanza, 2008; Kosny *et al.*, 2014). Interpreters can be professional or lay public members (e.g. friends or family). To ensure quality, I requested an interpreter from a professional interpreting organisation. However, using professional interpreters does not guarantee quality. Kosny *et al.* (2014) experienced situations when the interpreters did not act professionally, e.g. by summarising interviewees' answers rather than interpreting appropriately. This approach might be ethical in day-to-day interpreting but is a poor research practice. Training and briefing interpreters about their roles can overcome this challenge (Kosny *et al.*, 2014; Edwards, 1998; Sharp and Randhawa, 2017). Bringing a new person to research has challenges, especially around privacy. Confidentiality between the interpreter and interviewee is needed (Merry *et al.*, 2011). The interpreter and the participants did not know each other.

In this thesis, South Asian participants were asked if they wanted an interpreter, and one interviewee opted to have the conversation in Urdu. Also, for one of the public involvement groups, the discussion was in Polish, but as the artist attending the session who created the visual minutes[26] did not speak the language, the interpreter was on the phone supporting the artist in creating the visual minutes.

I now discuss how each of these interpreters was involved. Prior to the sessions, they received an interview guide and details about the study. This allowed them to familiarise themselves with specific vocabulary related to the research project and follow more easily the discussion. I also spoke with a Polish interpreter to clarify the focus group's practicalities. Interpreters joined zoom meetings 15 minutes before the participants and were briefed about the ground rules and their roles. This ensured that they did not become the creator of data but rather conduits of what participants said (including reporting any cultural nuance). At the start of the interview, interviewees were introduced to interpreters, and their role was explained, this time to participants. For the Polish group, the interpreter conducted whisper interpreting on the phone to the artist, whereas this was consecutive interpreting for an interview in Urdu. Interpreters also stayed after the interview for debriefing.

---

[26] See more information about this PPI session at page 147.

This subsection has shown that I undertook steps in this thesis to capture differences that might occur in research conducted in one language and reported in another. Cross-language research has challenges but also offers rewards if conducted appropriately. I already mentioned in this section collecting data remotely. This aspect of working online is discussed further in the next subsection.

6.3.1.1.2. Online interviewing

Textbooks on conducting interviews rarely considered online interviewing (often because, like the seminal work by Rubin and Rubin (2005), it was written before the rise of the digital age). However, the impact of the Covid-19 pandemic forced researchers to stop face-to-face work and move to online data collection. This subsection discusses the practicalities of online interviews.

Conducting interviews online has expanded during the Covid-19 pandemic (but was not a new method) as it offered a safe alternative to face-to-face meetings (Lobe, Morgan and Hoffman, 2020). In studies where researchers asked participants of Zoom interviews how their experience was, all (Gray *et al.*, 2020) or most of them (Archibald *et al.*, 2019) spoke positively about it. Online interviews have been shown not to influence participants' willingness to discuss sensitive issues (Sipes, Roberts and Mullan, 2022). Still, online interviews offer benefits and drawbacks to researchers (Archibald *et al.*, 2019; Deakin and Wakefield, 2014; Lo Iacono, Symonds and Brown, 2016). Online interviews are time efficient and reduce the cost of research as there is no need to travel for meetings (for both researchers and participants) and one can do it from the comfort of their home. On the other hand, there is a risk that technology could break and it would negatively influence the quality of the conversation and reduce potential rapport with participants. Qualitative research requires face-to-face interaction and seeing the non-verbal clues, this might be challenging during online interviews. Throughout the pandemic, many people worked and socialised on zoom, this could have led to zoom fatigue (Bailenson, 2021). Gray *et al.* (2020) recommended to not exceed one-hour session at the time to avoid fatigue. This was the time specified in the participant information sheet, and I tried to keep to it as much as possible.

Online interviews were not new for me as I conducted phone interviews in my previous roles as the study participants were spread around the country (including rural areas), and it was not feasible to travel to all of them. This provided me with the confidence to have a comfortable online interview. However, as these interviews were conducted online, I decided to follow established guidelines to ensure that all practicalities are covered (e.g. Sipes, Roberts and Mullan (2022) for Skype or Gray *et al.* (2020) for Zoom). I followed Gray *et al.* (2020) ten recommendations as they used the same technology.

1) Zoom was tested ahead of the interview, so I made sure that Zoom worked properly on the laptop and recording was of good quality when saved on the drive.
2) Providing technical support to participants, I checked with them if Zoom was the preferable option. No one said they were unfamiliar with it.
3) Consider a backup in the case of technical difficulties; if Zoom would not work, I planned to move to Teams, but this did not occur. However, sometimes I or a participant had broadband issues, so we switched off cameras to hear each other better.
4) Research should plan for distractions; I reassured participants it was completely fine if they needed to do any other activity and take a break. One participant asked if they could wash dishes during the interview as it would help them to focus.
5) Participants are provided in advance with a direct link to the scheduled Zoom meeting. Everyone received a unique link to their own interview.
6) I ensured that there was enough space on a hard drive to save any recordings. This storage space conformed with the ethics approval process.
7) and 8) recommendations are around ensuring a stable internet connection; I had fibre-quality broadband, so I decided that it was not necessary to hardwire the laptop to the Internet.
9) I included a reminder in the interview guide to turn on the recording.
10) All information about the study and consent forms were processed before the start of the interview.

In this section, I explored the qualitative interviews and examined the practicalities of conducting them in different languages, with the support of interpreters and online. The section has shown that qualitative interviews can offer a richness of data when following established practices. As qualitative, semi-structured interviews have been identified as an appropriate method for data collection, the next task to tackle is how to identify the right participants. This is the focus of the next section.

6.3.1.2. Sampling strategy

Sampling allows to identify and select appropriate participants for qualitative interviews. This section discusses how I sampled my participants.[27] In my PhD, I had to decide who to recruit for interviews. The research aimed to understand how to involve and engage seldom-heard communities, and the social constructionism perspective would argue for looking from multiple angles at this phenomenon. Thus, I decided to focus on seldom-heard communities and

---

[27] For participants demographics, please refer to later chapters, seldom-heard communities at page 197 and researchers at page 178.

researchers as two main participant groups. This offered to understand the involvement and engagement from the perspectives of involving/engaging and involved/engaged. Therefore, this section discusses these two groups. First, I focus on seldom-heard communities (Polish and South Asian participants). Second, I reflect on the choice of researchers. Thereafter, I justify the choice of purposive sampling.

As previously discussed, who is a part of seldom-heard communities can differ between projects[28]. It would not be possible to explore all of them within the lifetime of this doctoral research. I decided to focus on migrants and ethnic minorities as they are often recognised in almost any research as underrepresented communities. Although, further narrowing was needed as these communities remain diverse. My focus was on Polish and South Asian communities. I discuss in chapter 11 why these two groups are appropriate to study in the UK.[29]

Researchers working in big data offered the opportunity to understand how they have involved and engaged seldom-heard communities in their research. As the scoping review has shown[30], limited published literature has discussed this topic. Therefore, I chose to include their experience of the involvement and engagement of any seldom-heard community. This was a decision of my participants to decide if they considered their members of the public as seldom-heard communities. Some potential participants exchanged emails with me, discussing the inclusion criteria before deciding if they would like to participate in an interview. The category of researchers was further expanded to include public involvement and engagement facilitators; these include people responsible for and acting as a link between researchers and public members in big data research. Although for clarity, I refer to this group of participants as researchers.

The sampling for seldom-heard communities and researchers was purposive to select participants that allowed to address the research aims (Bryman, 2016) rather than to aim for generalisation. Seldom-heard participants had to live in the UK, be a part of Polish or South Asian communities and never previously be involved as public contributors in any research (not only big data). The last requirement was to ensure that the sample included interviewees who had not been previously involved as public contributors. Some research is available exploring the experiences of those already involved in PPI activities, but no study focused on

---

[28] See at page 26.
[29] See at page 197.
[30] See results chapter from the review at page 88.

people who have not been. This could be explained by the easiness of recruiting already committed public contributors.

Researchers had to work on big data research projects and involve or engage members of the public around it; they would have experience working with seldom-heard communities or planning to reach them. Keeping it broad between doing and planning was to encourage researchers who wanted to involve or engage seldom-heard communities, but they did not succeed. The researchers' sampling was initially purposive but then turned into snowball sampling as interviewees identified new potential participants (Bryman, 2016).

I accepted any potential participants who approached me as long as they met the above criteria. Therefore, I followed the rule first come first served basis. Therefore, the response rate was high as potential participants had to approach me in the first place. However, the recruitment of online participants has some risks as there has been the growth of fraudulent participants (Woolfall, 2023). This can be especially the case when participation is reimbursed. Only Polish and South Asian participants were reimbursed for their time, but the amount was not included in the public advert. I did not experience participants using poor quality audio or sending short emails when trying to sign up for the participation as Woolfall (2023) experienced. The lack of fraudulent participants could have been caused by the recruitment strategy as this was mostly through closed social media groups or shared by the organisations. In the case of researchers, all participants used institutional email addresses.

This section outlined the sampling of participants and presented the requirements for them to meet in order to take part in an interview. Based on that, the recruitment strategy was developed, which I discuss in the next following section.

6.3.2. Recruitment

This section discusses what participant recruitment looked like. The focus is on Polish and South Asian participants (as members of the public) and then on researchers (as professionals). It needs to be recognised that the Covid-19 pandemic and social distancing rules influenced how recruitment could take place. The ethics committee did not allow face-to-face contact with participants, so all recruitment had to occur online. It was possible to recruit study participants from even less technology-savvy groups (like older people) during the pandemic (Melis, Sala and Zaccaria, 2022). Social media was recognised as a successful space for participant recruitment in health research before the pandemic started (Whitaker, Stevelink and Fear, 2017).

6.3.2.1. Polish & South Asian communities

Social media (especially Facebook) enables researchers to reach people sharing specific characteristics (Brickman Bhutta, 2012; Baltar and Brunet, 2012) and has been previously used to reach migrants and ethnic minorities in the UK (e.g. for Polish communities, see Pötzschke and Braun (2017)). Recruitment for Polish and South Asian participants took place mostly on social media: Facebook, Twitter and WhatsApp. On Facebook, I joined groups for Polish or South Asian communities in the UK (e.g. Polish Diaspora in Merseyside (original in Polish: Polonia w Merseyside)). In Polish groups, these posts were published in Polish. If required by the terms of conditions of these groups, I approached admins or moderators, asking them for permission to post study ads. I also approached the organisations that I was aware of supporting migrants and ethnic minority organisations and was previously involved in (e.g. Polish Association Aberdeen) to share the recruitment call. Some of the interested participants emailed me, but many also directly messaged me on Facebook. As Facebook is about networking, users can invite each other to become friends (and thus have access to follow others' updates). Following the experience of Valdez *et al.* (2014), who had participants sending friends requests, I decided not to accept these from people interested in research. This was to ensure that my privacy was protected.

I initially also posted the recruitment ad on Twitter, and then I encouraged people to retweet it. Public contributors mostly used WhatsApp as they were members of some WhatsApp groups for South Asian communities and shared study details. Previously, researchers utilised paid advertisements to target specific participant groups on social media (Pötzschke, 2022). Due to a limited budget, I decided not to pursue this option as I preferred offering a higher voucher value for participants. Recruitment on Facebook was the most successful, and some participants also mentioned WhatsApp. There was no one speaking of learning about the study from Twitter. The reason behind it can be that Twitter remains mostly a professional space. This aligns with the experience of other research as identified in the systematic review by Whitaker, Stevelink and Fear (2017).

If not for the Covid-19 pandemic, more traditional ways of reaching participants from migrant and ethnic minority communities include posters and leaflets shared through ethnic minority organisations, religious places and public bodies. Despite that, Kayrouz *et al.* (2016) suggest that when recruiting seldom-heard communities, Facebook can be more efficient than traditional methods. The limitation of online recruitment was that people with limited computer skills might not have participated in the study. One of the South Asian charities said they had a group of people they work with within the community interested in participating in the study. However, they preferred in-person interviews and at the charity’s headquarters. The workers were willing to arrange the space for me to conduct these interviews, and there would be no additional research cost as they would not charge me for the room. This was a generous offer, but I had to refuse and explained that the ethics committee did not allow face-to-face interviews due to the Covid-19 pandemic. This was met with a little resentment as the government's restrictions were already relaxed at that time and people were allowed to meet in person (e.g. go to the pub).

6.3.2.2. Researchers

Recruitment of participants who were researchers was mostly undertaken in the professional sphere. Twitter was a part of it but not the main recruitment avenue. However, many Twitter users retweeted it (including the International PPI Network (@GlobalPPINet) and @usemydata, which is a patients' movement in the UK focusing on using patient data to improve healthcare). I approached the organisations that work daily with big data research and involve or engage members of the public around it. These included: Understanding Patient Data, the European Patients Forum, Public Benefit and Privacy Panel in Scotland, Health Data Research UK, NIHR ARCs, and Patient-Centred Outcomes Research Institute. The response was positive, and many organisations shared the recruitment call through their internal or external bulletins. I did not record where the participants learnt about the study, but some told me that they saw it in multiple places.

The Covid-19 pandemic did not influence the recruitment strategy for this group, as professional networks were primarily online before the pandemic. Second, as all participants' work was moved online, they were comfortable using Zoom for meetings, and some attended the interview from their home office.

This section has demonstrated that online recruitment was a safe and successful way of reaching potential participants during the Covid-19 pandemic and offered an opportunity to reach people around the UK and outside. Also, there was no financial cost as there was no need to travel and arrange space or refreshments (Whitaker, Stevelink and Fear, 2017).

6.3.3. Interview procedure

This section discusses the process of conducting interviews with Polish and South Asian communities, and researchers. First, I explore the development of topic guides and then pre, during, and post-interview procedures with participants. Public contributors were involved in designing and piloting qualitative interviews, and a detailed discussion of their role is covered in another chapter[31].

---

[31] For more details see chapter on public involvement in my thesis at page 147.

6.3.3.1. Topic guides

Two topic guides were developed to steer the conversation with participants, one for each group (see Appendix 2 for Polish and South Asian participants and Appendix 3 for researchers and PPIE facilitators). The initial drafts were based on the previous literature, but soon, I recognised that additional input was needed due to limited research on this topic. I showed and discussed guides with my supervisors, who provided suggestions, and a further iterative process took place with public contributors. The topic guide for interviews with Polish and South Asian participants was translated into Polish for those interviews when participants requested a conversation to be held in Polish.

The topic guides did not have to be strictly followed but provided me with the directions to ensure that all participants covered topics related to the research questions. The interview questions were broad and divided into different topical sections that aimed to construct the interviewees’ experiences. This structure was what Rubin and Rubin (2005) describe as a tree and branch. That structure ensured the breadth and richness of responses while also providing the interviewees' main questions. The interview started with a broader warm-up conversation and always finished with a wrap-up discussion. The warm-up questions aimed to ensure that interviewees received questions that they could answer easily, build their confidence and establish rapport between myself and the participants. To support the collection of rich data, questions included multiple prompts and follow-up questions, thus signalling the interviewees' need for depth in their responses (Rubin and Rubin, 2005). The wrap-up sections signalled that the interview was coming to an end and offered participants an opportunity to add anything else to what was already discussed. Thus, participants could have added more information to previous questions or suggested other issues not covered by myself. The topic guides provided a reminder of issues to mention before the interview started. These consisted of reiterating key points from the consent form and participant information sheets, such as ensuring they received both of them and are happy with their contents, reminding them that session will take around 60 minutes (Polish and South Asians participants) or 30 minutes (researchers and PPIE facilitators), possibility to take breaks if required, reassurance that they did not have to answer any question and there were no right or wrong answers as the aim of the conversation was to understand their views and perceptions.

Big data research is often an unfamiliar topic to members of the public. Therefore, other researchers, such as Bailey Wilson et al. (2022), provided participants with an explanation of how patient data can be reused for research. The topic guide included the lay summary of big data research, which was shared with each participant at the beginning of the interview. In addition, there was a link to the Data Saves Lives campaign website with lay information providing examples (Data Saves Lives, 2023). Professional participants’ (researchers and the PPIE facilitators) guide also had definitions of public involvement and engagement, as these can differ between regions (Lauzon-Schnittka *et al.*, 2022). This ensured clarity when the participants and I referred to involvement and engagement.

6.3.3.2. Prior to interviews

Interested participants had to reach me through email requesting more information about the study. They received the participant information sheet (see Appendix 4 for a copy of the information sheet shared with Polish and South Asians participants), and I encouraged them to ask any questions or respond if they would like to participate in the interview. The consent form was sent through email, and participants were requested to fill it out and send it back. One participant had technology problems and could only get access on their phone, so their consent was recorded in a separate file.

Basic demographic information was collected from my professional participants (researchers and the PPIE group) around their length of experience in research or PPIE, their role (if they are a researcher, PPIE facilitator or both) and the country they are based in. For Polish and South Asian participants, only their ethnicity was recorded throughout the interview and if they were at least 18 years old.

All interviews took place on Zoom as it offered privacy, and only participants I allowed to join could participate (Zoom, 2021). This ensured that no one except participants had access to the interview. Zoom also does not require an account or any installation for people to access it. Therefore, I decided it would be the easiest option for people to use (e.g. in contrast to Skype, which requires an individual account). It was important to ensure that participants did not have to spend much time preparing before the interview. However, I still noticed that some joined the link a day or two before to test if it worked (Zoom emailed me that someone was already in the waiting room). This is similar to what Melis, Sala and Zaccaria (2022) experienced: technology-anxious participants tend to log in earlier to see if everything works.

6.3.3.3. During interviews

When reminding people about the time commitment and the expected length of the interview, the majority of participants were happy with it. Still, some, especially in the researchers' group, pointed out that they had to attend another meeting straight afterwards, so they had to finish sharp on time. In the middle of the interview, others recognised that they wanted to share more and asked if it was okay to speak longer. I always accommodated interviewees.

Participants were in different locations and used phones or computers. I always logged in to the institutional Zoom account on the university laptop.

During interviews with researchers and PPIE facilitators, they often anticipated the following questions and answers. In that case, I skipped that question. This aligns with Rubin and Rubin (2005) experience that interviews sometimes can follow a bit different path.

Traditional face-to-face interviews should take place in a quiet and confidential location. I did not have an influence on where my participants were during the Zoom call. Some participants were in their homes, offices or gardens. On a couple of occasions, they requested a break to look after kids or home deliveries.

6.3.3.4. After the interview

As the interview concluded, I thanked the participants and explained the process of receiving reimbursement, offering them a choice of Amazon or a high street shop voucher. Then, I asked all participants if they would like to receive a summary of the findings. All researchers request it, and the majority of Polish and South Asian participants.

For many Polish and South Asian participants, the concept of being public contributors was novel, but they were keen to learn more and potentially become one. Many participants asked where they could read more; if they did not, I shared it myself. When requested, I send them contact details to their local PPI groups, usually within the NIHR network. A minority of participants were also asked for additional information to learn more about big data research in their spare time, and I provided them with some further reading. Information was written in a lay language and targeted public members, developed through the Data Saves Lives campaign. These materials were available only in English.

The Zoom saved the recording on the local disk of the university-provided, password-protected computer. Then, it was moved to the secured university server.

After each interview, I took notes for reflection and to ensure the quality of the research. This allowed me to reflect on what participants said and if they understood the questions in the intended way. When the interpreter was present, I asked them to stay after the interview for a debriefing to check if they felt that everything was interpreted correctly and with enough detail. Also, I asked them if there were any cultural nuances which could have been lost in the translation. These comments were included in my post-interview notes.

Recorded data needs to be transcribed later for analysis. The process requires detailed work and can be considered a first step for successful analysis (Bailey, 2008). A researcher should decide if visual data (e.g. hand gestures) and emotions would be relevant to capture. Technology can facilitate the process of transcription. I transcribed all interviews myself with the support of Microsoft Teams. The software automatically transcribed the conversation, and

I relistened it twice to check the transcript quality. No visual data was recorded in the transcripts, but if participants had an emotional response (e.g. they laughed), I recorded it (manually) in the transcription.

This section has offered an overview of the interview procedure. I examined the topic guide for semi-structured interviews and reflected on how I engaged with participants prior to, during and after the interview.

6.3.4. Twitter data

The second source of data was Twitter. This section briefly discusses research through social media, especially Twitter. I discuss the practicalities of data collection. Then, I recognise some limitations of this kind of research and reflect on possible ways to overcome them.

Modern technology has led to social media's growth, allowing researchers to creatively reuse that data for research (Kara, 2020) even if social media users did not post it initially for research purposes. Multiple social media platforms exist, with Twitter being a leading, free-to-use microblogging platform (Twitter, 2022). Twitter users can post short online posts (called tweets) of up to 280 characters to engage with others. The platform is interactive, as users can respond to others' messages, retweet or like them. Everyone can read tweets, but only registered users can post new ones. There are around 238 million Twitter users worldwide; therefore, it has become a popular space for researchers and policymakers to engage with colleagues and members of the public (López-Goñi and Sánchez-Angulo, 2018; Fuller and Allen, 2016; Tripathy *et al.*, 2017).

Initially, a researcher needs to identify a hashtag that allows them to read all tweets around this topic. This can be challenging as these can change as Twitter users can create new ones anytime, and these can become viral. This requires the researcher to add a new one potentially. Teodorowski (2020b) initially used one hashtag, but before the studied event took place, another two relevant hashtags became popular on Twitter and were consequently added to the dataset. In my research, this was easy as the hashtag was well-established and in use for the last couple of years (Hassan, Nenadic and Tully, 2021).

Twitter data needs to be downloaded from social media for analysis. This can be conducted by one of the available programmes, such as Tweetcatcher or NCapture. Previously, researchers successfully used them for qualitative research on Twitter (see Talbot *et al.* (2020) for Tweetcatcher and Hays and Daker-White (2015) for NCapture). As I wanted to conduct the qualitative analysis supported by NVivo, I used NCapture as both applications are compatible. NCapture is an extension to a web browser (in my case, Chrome) (NVivo, 2022). It saves all downloaded tweets into one dataset directly to NVivo. Only publicly available tweets are downloaded. In other words, these Twitter users who restricted access to their posts would not be accessible. Downloading tweets through NCapture requires the researcher to have a Twitter account. When logged in to Twitter (using my personal account @PTeodorowski), I chose the content to download in the app. I ensured that tweets were downloaded as dataset and named each file. Within that dataset, the following information is recorded: username, full tweet (but without emoticons and pictures), time, tweet type (retweet or original tweet), number of retweets, location, bio of tweeter user, total number of tweets by the user, number of their followers, and how many others they follow. NVivo can assist to organise these data in descriptive statistics or visualise on the map. NVivo organises the datasets and automatically removes duplicates. However, the dataset might not be ready for analysis immediately and require further checks. This might include double checking for duplicates, tweets that include only hashtags or unrelated messages (e.g. spam), written in other languages. This is a manual process that needs to take place before the analysis.

Social media research is a new method and has some limitations. It is possible to search for all tweets using the hashtag systematically. However, if someone made a typo in the hashtag, these would not appear in the dataset. There is no guarantee that individuals write tweets as they could be created by bots (a form of software that controls a Twitter account and produces content on its behalf). These can automatically post new tweets, retweet, like others' content, and follow or unfollow other Twitter users. I was careful when analysing tweets to look for this kind of activity. I noticed some spam tweets removed from the datasets but did not notice any bot activity. However, this does not mean that there was not any.

This section has presented Twitter research as a novel way of collecting rich qualitative data. I explained a step-by-step process of collecting tweets for analysis. Thereafter, I acknowledged the limitations and dangers of conducting research on Twitter.

6.3.5. Data saturation

The previous sections discussed the data collected through qualitative interviews and on Twitter. There is a danger that an inappropriate amount of data would lead a qualitative researcher to “cherry picking” later in the analysis phase (Morse, 2010). This could produce inaccurate or even misleading findings. Therefore, the next issue to consider is the question of how many interviews or tweets are enough for the research, which this section focuses on. First, I briefly examine the importance of numbers in qualitative research. Then, I explore the concept of data saturation with some practical considerations. Third, and based on this discussion, I justify the saturation for each of my participant groups.

Qualitative research is not about numbers (or large samples), but Sandelowski (1995) points out that numbers when determining the right sample for qualitative research are not unimportant. This is to ensure the richness of data but also to avoid being "*drown[ed] in data*" (Morse, 2000, p. 3). Vasileiou *et al.* (2018) conducted a systematic review of qualitative health research papers to explore how authors discuss sample sizes. They found that the majority of papers in their sample did not provide any sample justification. Out of those that provided these details, most invoked the principle of saturation. However, saturation could mean different things and the concept has multiple uses. The second most popular justification was pragmatic considerations around challenges to access participants or time limitations. A recent systematic review by Hennink and Kaiser (2022) of studies assessing saturation in qualitative research found that participant groups might require only nine to seventeen interviews and in all circumstances data saturation was reached for in-depth interviews under 25 conversations. However, authors recognise that their results should not be used as guidance when deciding on the sample size in qualitative research.

Data saturation is often associated with grounded theory (Corbin and Strauss, 1988). It means that new data does not contribute new theoretical insights towards the research question (Corbin and Strauss, 1988; Bryman, 2016). In other words, there is an information redundancy (Lincoln and Guba, 1985) as no new codes or themes are extracted from the data. Saturation was often described as the golden standard in deciding when there were enough interviews. However, the data saturation has received some criticism from qualitative researchers (Braun and Clarke, 2021c; Braun and Clarke, 2021d; Malterud, Siersma and Guassora, 2015). Malterud, Siersma and Guassora (2015) prefer to use 'information power'; the richer in information the sample is (the more appropriate data in relation to the research it consists of), the lower number of interviewees is required.

Guest, Bunce and Johnson (2006) conducted an experiment with sixty qualitative interviews to understand the process of data saturation. They aimed to know when during the analysis, no new theoretical insights would appear in the next analysed interview. It took them to analyse twelve interviews to cover 92 per cent of the generated codes. Their experiment has shown that even very small samples can be appropriate for qualitative research to offer a richness of data. However, one of their research limitations was that the interviewees were a homogenous group that discussed a narrow topic.

Qualitative researchers should use their judgment to justify the sample size (Sandelowski, 1995). It might be appropriate to initially overestimate the sample to ensure appropriate timing and resources (Morse, 2000). The final number of interviewees cannot be decided in advance of the analysis (Braun and Clarke, 2021d). My decision has been guided by multiple factors such as data quality (and its usefulness obtained from interviewees), the aim of the study, the study design and the choice of qualitative method (Morse, 2000). This overlaps with the concept of ‘information power’, but I continue to use data saturation as it remains the most utilised among qualitative researchers.

Based on this discussion, in table 3 I justify the final decision for each of the participants' groups.

Participants	Reflection on data saturation
Polish and South Asian participants	Data saturation was reached after twenty interviews with each ethnic group, and no new information relevant to research questions was appearing during interviews.
Researchers	Twelve interviews were conducted with researchers (including PPIE facilitators) as no novel data around techniques of involving and engaging seldom-heard communities in their work appeared.
Twitter data	As qualitative research on Twitter is still developing, I supported my decision on data saturation based on similar research projects utilising the same analytic methods in their research. Many qualitative Twitter studies successfully conducted the qualitative analysis with less than 1000 tweets (Berry *et al.*, 2017; Bogen *et al.*, 2022; Rashid, Mckechnie and Gill, 2018). This approach has been utilised successfully by Vasileiou *et al.* (2018).

Table 3 Summary of data saturation for each study

This section provided a discussion on data saturation. Using this discussion, I also provided the justification for how I knew when to stop the data collection in my studies.

6.4. Reflexivity & positionality

Reflexivity has multiple meanings for researchers (Bryman, 2016; Lynch, 2000). This thesis perceives it as researchers’ reflections on the implications of their values, biases and personal experiences in constructing knowledge and choosing methods. As much as many qualitative researchers recognise reflexivity as the assurance of quality and robust research, it has received criticism. Lynch (2000) raises questions if a reflective researcher really is somewhat superior to an unreflective one. Before modern reflexive practice, researchers (in sociology and anthropology) had their own reflective approaches, even if they were not explicitly called reflexivity (Lynch, 2000). He further argues that due to multiple meanings of reflexivity,

reflexivity as an act remains insignificant unless it is embedded in a relevant theoretical and contextual application. To overcome that challenge, my reflections are based on the social constructionism perspective that reflexivity means explicitly acknowledging my personal values and perspectives that might influence the research (although I recognise that this is one of the meanings social constructionism might give to reflexivity) (Burr, 2015). This approach is consistent with Braun and Clarke (2021c) reflexive thematic analysis that argues for positioning oneself (ethnic background, age etc.) in relation to analysed data. Understanding our biases helps us to see what we bring to the analysis. Reflection is an ongoing process that starts even before research begins and is never final and complete (Braun and Clarke, 2021c). My reflections took place throughout this thesis, often recorded in writing so I could return to them again, and I also involved public contributors in reflection tasks.[32]

This section reflects on my positionality. Following Wilkinson (1988), I discuss three types of reflexivity: personal, functional and disciplinary. Personal reflexivity focuses on the influence of my values. Functional reflexivity explores how my choice of methods and study design impacted the construction of knowledge. As these two are interlinked, they should be discussed together. Disciplinary reflexivity is about how my discipline shapes research.

6.4.1. Personal and functional reflexivity

Reflexivity is '*the me-search within re-search*' (Burnard as cited in Kara (2020, p. 95)), so I discuss my personal biography that had an impact on the research. This includes my nationality (being a migrant) and professional background. This discussion is based on the migration literature as two participant groups, Polish and South Asians participants, are from migrant or ethnic minority communities in the UK. Then, I discuss how my professional and academic trajectory influenced the choice of this research topic.

In migration studies, the positionality divide has traditionally been between insider and outsider perspectives. An insider is a researcher from a migrant community who studies it. An outsider usually is a member of the general population (or is from another migrant background) and studies a migrant community. Involving insider and outsider perspectives among different research team members offers more robust data analysis (Teodorowski *et al.*, 2021b). However, the insider-outsider divide has been challenged as it is much more complex (Carling, Erdal and Ezzati, 2014; Voloder and Kirpitchenko, 2016). A migrant researcher can be both an insider and an outsider in their own community. They speak a language and understand the culture and experiences of the community, so it is seen as one of them. However, this

---

[32] See discussion on positionality of public contributors at page 150.

attribute does not exclude having an outsider status. There are other attributes that could cause a feeling of being an outsider among people coming from the same ethnic background, such as age, education, class or different integration trajectory (e.g. depending on language skills) (Carling, Erdal and Ezzati, 2014).

Shared ethnic backgrounds with participants can open new avenues of inquiry. Mapedzahama and Kwansah-Aidoo (2016, p. 169) write of the participant who said it was ‘refreshing’ to speak directly with people understanding their experience of being a black person in Australia. My experience was similar to theirs. Some of the Polish participants said that they decided to participate in interviews because I am Polish (and thus they wanted to help another Polish person), because it was in their mother tongue (otherwise, they would not be comfortable speaking in English and have mixed experiences of interpreter support) and felt that I could really understand them (as was on the same migrant trajectory as them). If I was not a researcher from a Polish background, some of my Polish participants would not have participated in an interview or might not have been so open when discussing their experience. Although this also meant that interviewees assumed that due to sharing my mother tongue, I shared the same migration trajectory with them (Mapedzahama and Kwansah-Aidoo, 2016; Gawlewicz, 2016). However, as I am a part of the middle class and completed higher education in the UK, this would not have always been the case.

Being an insider can put some additional pressure on the researcher. Andits (2016) studied the Hungarian diaspora in Australia as ethnic Hungarian. The researcher's status gave her authority in the community (in addition to being a young person), but it felt that some participants pressured her to ensure that younger generations would be more involved in the diaspora life. During interviews with Polish participants, interviewees spoke about the barrier between researchers and the Polish community. When asked what can be done to overcome it, some participants suggested that my role as a researcher from the Polish community is to become that bridge between communities and solve that challenge.

The second participant group were from South Asian backgrounds. Thus, based on the insider-outsider divide, I assumed that I would be perceived as an outsider. To overcome this challenge, I involved two public contributors in my research who are from South Asian backgrounds. My intention was that their involvement would help me to reach and build trust with South Asian participants. As a white Polish man, I was concerned that European (especially British) colonial histories might affect my relationships with South Asian participants as many of them came, or their ancestors originated in former British colonies. However, this was not the case as participants saw me as one of the migrants, even referring to me as a first-generation migrant in the UK. They assumed that my living experience in the

UK was closer to theirs than the British population. Often, I was asked how long I had been living in the UK. This eased initial contact and building relationship with potential participants before and during interviews. My experience would fit what Carling, Erdal and Ezzati (2014) describe as ‘insider by proxy’ when a researcher from migrant background studies another migrant community. That common migrant status established a sense of commonality between the researcher and the participant.

Professional background can influence the researcher's perspective. My professional background is diverse. In addition to academia, I worked in a charity focusing on the integration of ethnic minorities, where I was responsible for involvement and engagement, mostly for projects around NHS services. This gave me a broad understanding of how patients and members of the public (especially from ethnic minorities) can shape their local services, and the barriers they experience. I saw first-hand how healthcare staff or researchers respond to members of the public views and suggestions. I build on this experience by volunteering with Polish charities and actively supporting the diaspora. Mcgrath, Palmgren and Liljedahl (2019) recommend qualitative researchers with clinical experience to use their knowledge of the healthcare service during the interviews. Thus, they could explore the topic more in-depth than the interviewee without a clinical background. I believe this applies to researchers with other backgrounds too. My journey with public involvement and engagement started in 2015, and I used this knowledge to probe participants' experiences of how researchers could reach them. The way how I collected data could have looked significantly different than by a person who had never involved or engaged seldom-heard communities.

6.4.2. Disciplinary reflexivity

My research background is interdisciplinary, as I studied various disciplines in sociology, politics, legal studies and international relations during my undergraduate and postgraduate years. My interest in seldom-heard communities started during my postgraduate degree when I explored the integration of Syrian refugees in the local area (Teodorowski, 2020a). My expertise started to cover PPIE when I worked for a local charity. My journey with both topics started in 2015 and overlapped with different, primarily policy-related projects over the following five years.

Using interdisciplinary lenses is helpful in PPIE as this field is still developing and remains used among health professionals and non-clinical researchers. Social constructionism encouraged me to consider multiple perspectives equally (Gergen, 2015). This differs from other paradigms, which can prioritise one view over another. For example, positivism devalues the data coming from qualitative research (e.g. interviews) as it considers it preliminary (Wilkinson, 1988).

Choosing a topic to study when you are not an expert in it can be challenging. I am not a big data researcher and I was concerned at the beginning of my PhD that I would not understand the complexity of this research area. However, this was not the case, as I was involved in the work of a cross-thematic group in my funder's organisation. Through my interdisciplinary lenses, I brought expertise from other fields into PPIE in big data research.

This section has examined how reflexivity allowed me to locate myself within my own research and encouraged creative thinking (Kara, 2020). I took time from my research to understand how I know what I know and why this is the case. Understanding my own positionality benefited the analysis and interpretation of qualitative data. Although the reflections were not limited to my experiences and in the public involvement section, I discuss how the reflexivity with public contributors benefited the analysis further.[33]

6.5. Ethical considerations

Qualitative researchers need to consider ethical issues arising from studying human participants. Some of the ethical problems include voluntary participation, informed consent, confidentiality, potential benefits and risks, and secure data storage. In this section, I discuss each of these issues. Here, also the influence of the Covid-19 pandemic on ethics are recognised.

6.5.1. Ethical approval

Following university guidance and ethical standards, ethical approval is required in academic research (Bryman, 2016). These rules exist to protect participants, researchers and the university. Before the commencement of the research projects, the Liverpool University Ethics Committee approved all studies under references: 9815 and 10063. Appendix 5 and 6 present the approval letters from the ethics committee. All documents related to the research projects, such as participants' information sheets, consent forms, interview guides and ads, had to be attached. The ethics process was smooth, with only minor revisions suggested by the review panel. However, I experienced what Lincoln and Tierney (2004) describe as ethics panel members sometimes not understanding how data will be generated or used in qualitative research. Their experience relates to my ethics application for the Twitter study. At the University of Liverpool, all ethics applications have to be submitted through an online form, but there was no separate form for social media research at the time of my application. After speaking to ethics officer, I was asked to use option ‘other’. This was confusing as many text boxes seemed to require the same (or similar information), and I often duplicated the responses, trying to answer overlapping questions. There was no negative feedback on how

---

[33] See at page 150.

the application was written. However, the ethics officer told me they were working to create a special form for social media research.

6.5.2. Voluntary participation & informed consent

All participation in the interviews was voluntary and undertaken after informed consent had been given by participants. This principle is clear but challenging to implement as Bryman (2016, p. 129) points out that "*it is extremely difficult to present prospective participants absolutely all the information that might be required for them to make an informed decision*". To overcome this challenge, I used standard forms developed by the University of Liverpool and involved public contributors in the process of developing them to ensure all relevant information was provided in an accessible language. Written consent forms became a standard in research (Bryman, 2016). They ensure that participants are provided with full information about the study and that researchers are protected if any issues are raised around the process. Often, these would be obtained in person, with the researcher keeping one copy and the participant the second. However, this was not possible because of Covid-19 restrictions. This caused difficulty because I did not know if the person who digitally signed it was actually the participant. Following the experience of other researchers, I followed two step consent process (Thunberg and Arnell, 2022). First, as participants said that they would like to take part in an interview, they received through email the consent form to sign digitally and send it back to me before the interview. They were also encouraged to ask any questions. I checked if the form was completed correctly and, if not, sent it back to the participants. This happened, for example, when the participant attached an unsigned version. Second, at the start of the interview, I also asked if they had any questions and confirmed it was actually the participant who signed it.

Obtaining informed consent from social media users is more challenging. On the one hand, all data online is publicly available (unless users choose privacy settings not to allow it). Twitter specifies that tweets can be quoted but need to keep the original text and include the user name. Thus, from that legal perspective, there is no need to get consent from social media (such as Twitter) users if the researcher wants to reuse their tweets for the analysis. On the other hand, researchers should protect participants' privacy (Bryman, 2016), for example, by not collecting identifiable information that can be linked with individual responses. This can be hard to ensure because direct quotes, even when anonymised, can be easily identified through any search engine. There has been a mixed approach in published papers around this challenge. Researchers approached authors of direct quotes for their consent (Talbot *et al.*, 2020), recommended paraphrasing (Townsend and Wallace, 2016) and anonymised tweets (Hays and Daker-White, 2015).

UK Research Integrity Office (2016) recognises that there is no ‘one-size-fits-all’ guidance on internet research as the area is still developing. My approach was inspired by evidence of what Twitter users think about this topic and staying within the legal framework of Twitter. Williams, Burnap and Sloan (2017) conducted a study exploring Twitter users' views towards reusing their data in social research. They found that most of their participants were not concerned about their social media data being reused for academic research; however, they also expected to be asked for consent before publication. In accordance with it, I decided to approach Twitter users (excluding public bodies and organisations) to inform them about it and give an option to opt out. None of them told me that they want to opt out but one asked for a copy of published paper (which was provided afterwards).

6.5.3. Confidentiality

I guaranteed the interviewees anonymity and confidentiality. Thus, their identity remained unknown to others. All participants were assigned pseudonyms, and if they provided any recognisable information about their background, work or where they live, I altered it in the transcript and, if necessary, further in quotes. This was especially a challenge when anonymising interviews with researchers as they were asked to describe their research or involvement projects to provide a context to the discussion. One of the participants (who was a researcher) asked me what was put in place to ensure that what they say would remain anonymous and confidential. I reassured them about confidentiality before the interview and explained the process. Online interviews might offer additional anonymity as participants might decide not to turn on their cameras and use only audio (Thunberg and Arnell, 2022); this was especially the case for some Polish and South Asian participants.

6.5.4. Potential benefits & risks

All research can have potential benefits and risks. In the participant information sheet, I specified that there are no direct benefits for participants for taking part in the interview. Still, the project aims to help researchers to better understand how to involve and engage with seldom-heard communities in big data research, so better involvement and engagement could be an indirect benefit. Participants were offered a £20 shopping voucher for the interview (Polish and South Asian participants). Providing financial reimbursement has become popular among qualitative researchers, but little guidance exists around it (Polacsek, Boardman and McCann, 2017). Reimbursement recognises participants' time and compensates for any expenses they could have incurred. However, some criticism has been that payment can be coercive as they induce participants to be involved in the study (Largent and Fernandez Lynch, 2017). Therefore, there was no information about the value of the voucher in the study advertisement, but it was specified in the participant information sheet. The decision around the amount was influenced by the public contributors' opinion and recognised the growing cost of living costs. The ethics committee approved the amount offered to participants.

I did not expect that the discussion topic would be distressing to participants. However, I offered an opportunity for people to refuse to answer any question, terminate the interview or take a break without giving any reason. None of the participants refused to answer any questions or stopped the conversation. However, some of them took a break, e.g. to check on a child or pick up a takeaway or delivery.

Conducting research has potential risks for researchers, and its impact requires a holistic approach to support researchers' well-being (Kara, 2018). This consists of physical safety, and emotional and mental well-being. Conducting online interviews has a limited risk to the researcher. This was especially the case in my research as the topic was non-sensitive. Despite that, in the ethics application, I stated that if I felt at risk from participants, I would terminate the interview and discuss my concerns with the supervisor. This was not the case. Also, I held regular meetings with the supervisory team that allowed me to discuss my experiences. Emotionally, research can be time-consuming and exhausting , e.g. conducting interviews, data analysis (Kara, 2018) or seeking ethics approval (Monaghan, O'Dwyer and Gabe, 2013). In addition to regularly meeting my supervisors, I attended peer support sessions with other PhD students, discussed our common experiences, and supported each other. PhD research can be very stressful, and research has shown that PhD students experience higher stress levels than the rest of the population (Hazell *et al.*, 2020). Covid-19 lockdowns further increased depression, anxiety and stress among PhD students (Paucsik *et al.*, 2022). Engaging in self-care is beneficial for students to alienate stress (Hazell *et al.*, 2020). I often took breaks to walk or exercise to improve my mental well-being. This allowed me to keep a work-life balance.

6.5.5. Data storage

Securely storing participants' data is important. All files were downloaded on a password-protected laptop provided by the University of Liverpool and stored on the secure University M-drive to which the supervisor and I had access. Audio recordings of interviews were deleted after transcription.

This section has presented key ethical issues in qualitative research. I discussed how I approached them and what was put in place to protect participants and myself as a researcher.

6.6. Conclusions

This chapter started with setting up social constructionism as the philosophical stance taken in this thesis. That choice influenced the use of qualitative methodology, which was discussed.

Then, I examined two appropriate qualitative research methods to answer research questions: qualitative interviews and Twitter data. The shift (due to the Covid-19 pandemic) to conducting interviews remotely and collecting data in different languages was examined. Then, I provided a detailed process of the interview procedure. After that, I reflected on my positionality as the researcher and considered how this could impact how I conduct the research presented in this thesis. Last, I considered the ethical principles relevant to qualitative research and explained how I ensured that my research was conducted ethically and in line with the requirements of good practice as set out by the University ethics committee. The next chapter follows these discussions as I present how I analysed the collected data.

7. Data analysis

7.1. Introduction

This chapter focuses on how I analysed my research data. As Braun and Clarke (2021a) argue, there is no perfect analytical approach to qualitative research. They do not think that their reflexive thematic analysis is the best or only approach, but suggest considering different analytic methods and methodologies that could be appropriate to different research questions. Therefore, first, I explore the choice of reflexive thematic analysis as the most appropriate analytical method to support my analysis. Second, I discuss the six steps of that method and reflect on how I followed and applied them in my work. Third, I explore my orientation to data. Then, the limitations and misconceptions of reflexive thematic analysis as an analytic method ware discussed.[34] Last, I provide some final reflections on the process and my experience with this analytic method.

7.2. Reflexive thematic analysis

There exist multiple qualitative analysis methods. In this section, I briefly outline my choices and justify why the reflexive thematic analysis was the most appropriate for this thesis.

Alongside thematic analysis as a qualitative analytic method, there are interpretative phenomenological analysis (Smith and Fieldsend, 2021), grounded theory (Strauss and Corbin, 1997; Strauss, 1990), discourse analysis (Wiggins, 2016) and many others.[35] These are structured approaches and might come with a pre-prepared package of theory (e.g. grounded theory), analytic method and research design (Braun and Clarke, 2021a). On the other hand, thematic analysis is more flexible as it does not consist of any pre-package theoretical commitments. This offers a researcher more flexibility but also comes with challenges as researchers often inappropriately apply it in practice (Braun and Clarke, 2023).[36] Reflexive thematic analysis should not be conducted without a theoretical underpinning and it is not a descriptive approach. I perceive this flexibility as a strength that qualitative researchers can effectively apply in the analysis.

Another qualitative analytic method is a qualitative content analysis which could be considered one of the methods most similar to thematic analysis (Vaismoradi, Turunen and Bondas, 2013). Qualitative content analysis is also a broad family of different approaches (Hsieh and Shannon, 2005). As these are often related to quantitative content analysis, researchers using qualitative content analysis tend to position themselves within postpositivism (as they try to

---

[34] I discuss data limitations in the discussion chapter at page 231.

[35] Each of these analytical approaches has rich diversity of how these are applied so I provided only some references as examples of how these are used in qualitative research.

[36] See more on the common mistakes when using reflexive thematic analysis in the section on strengths and limitations at page 140.

limit the researcher's bias by limiting subjectivity, for example, by supporting multiple coders). Consequently, studies using qualitative content analysis tend to be more descriptive (Vaismoradi, Turunen and Bondas, 2013). Braun and Clarke (2021a, p. 40) go a step further and call it "*accessible for qualitative beginners, but relatively unsophisticated*." As the research aims of this thesis were to provide critical, nuanced insights into the phenomenon, the qualitative content analysis was deemed inappropriate to achieve it. Therefore, thematic analysis approaches were deemed to be the most appropriate option.

Many researchers perceive thematic analysis as a singular method with clear procedures (Braun and Clarke, 2016; Clarke and Braun, 2018). For example, Vaismoradi, Turunen and Bondas (2013) compare content analysis and thematic analysis in their paper, but they do not specify any one approach to thematic analysis. Thus, they implicitly suggest that there is a 'one' approach to thematic analysis. However, thematic analysis is more of a family of different approaches. These include coding reliability approaches (Boyatzis, 1998), codebook approaches (King, Horrocks and Brooks, 2018; Ritchie *et al.*, 1994) and reflexive approaches (Braun and Clarke, 2006; Braun and Clarke, 2021c). Coding reliability approaches require multiple coders to reduce the researcher's bias when analysing data. They often use a more structured approach to coding and the themes coming from the data exist more as summaries of what dataset (e.g. participants in interviews) covered. Codebook approaches similarly use pre-determined codes. However, all coders create these as an analytical framework (Gale *et al.*, 2013). In other words, it becomes a chart for the analysis to be used by coders. Researchers would not necessarily code the same data. Reflexive approaches are based on researchers' skills, reflections and the need to recognise one's subjective perspective. The researcher's perspective is not a disadvantage to the analytic process but a vital resource that needs to be utilised (Braun and Clarke, 2019). Researchers identify codes in relations to the research questions and actively generate themes from them. This process is more organic, analytical and subjective.

Braun and Clarke (2021a; 2021c), following Kidder and Fine (1987) work, divide these thematic analysis approaches into "small q" and "big Q" of qualitative research. "Small q" refers to qualitative research tools or techniques that are not used as a qualitative methodology but rather orientated towards more quantitative research. On the other hand, "big Q" embraces qualitative research by embracing qualitative values. For example, the "big Q" approach would recognise that realities around us (and perspectives) can be multiple (this is a cross-over with the social constructionism perspective on multiple realities). Only reflective approaches fully follow the "big Q" approach. This thesis was embedded within qualitative research, so its analytic method should have aligned with the "big Q" approach.

All of these approaches to thematic analysis are well-used in health research, but the reflexive thematic analysis was the most appropriate for my PhD research. Reflexive thematic analysis is theoretically flexible and, according to Braun and Clarke (2021c; 2021a), works well with social constructionism, especially the version advocated by Gergen (2015), whose works I extensively refer to in the methodology section. In contrast, coding reliability or codebook approaches are more restrictive around theoretical underpinnings that could be used in the analysis. Coding reliability fits more within postpositivist perspectives, and codebook approaches are applicable to social constructionism but only when using discourse analysis as a research method. However, as discourse focuses on language, it would not be an appropriate method to address this thesis' research questions. Social constructionism argues for the recognition of subjectivity and biases. This is one of the central concepts in reflexive thematic analysis. There were also practical reasons for the adoption of this approach. First, it allowed me as a single researcher to code all data and include public contributors and supervisors at different stages; rather than jointly coding and checking agreements between coders (like in the approach suggested by Boyatzis (1998)). Second, my previous research used reflexive thematic analysis (Teodorowski *et al.*, 2021b; Teodorowski, 2020b), so I was familiar with this analytic method.

7.3. Conducting reflexive thematic analysis

Reflexive thematic analysis consists of six phrases (Braun and Clarke, 2006; Braun and Clarke, 2021c). The process is not always linear, despite it consisting of logical, sequential steps; however, these are flexible, and a researcher can move between them as needed taking an iterative approach. These steps include 1) familiarizing oneself with the data, 2) generating codes, 3) constructing themes, 4) reviewing potential themes, 5) defining and naming themes, and 6) producing the report (Braun and Clarke, 2006). I describe each of these steps and present what analysis I completed at each of them.

7.3.1. Familiarizing oneself with the data

The familiarisation stage consists of three practices. First, a researcher should build a detailed knowledge and understanding of the data they will be working with. Second, they should start looking at the text as beyond a form of information but rather as data that needs to be critically analysed. These two practices might feel at first sight incompatible with each other (Braun and Clarke, 2021c). The former asks for immersion in the data, but the latter asks for more distance and a critical approach. However, they are both essential for qualitative researchers. The third practice is to ensure that in-depth notes accompany these two processes.

Immersion in the data is easier for those researchers who collected data themselves. In my case, it differed between projects. I conducted interviews myself, so I was familiar with the conversations. Also, I transcribed all interviews myself. However, I only downloaded Twitter data. Therefore, it took me additional time for the Twitter study to familiarise myself with the dataset. For example, I read how the discussion developed on Twitter every couple of days and then read the whole dataset multiple times before moving to the next stage. Similarly to Byrne (2022), I took notes of potentially interesting data segments and noted my initial ideas regarding the dataset. These were collected in OneNote as the programme allowed me to organise notes in different categories. There were also some brief written notes made when I conducted the interviews and I often made notes straight afterwards to capture my immediate thoughts and impressions, and I included these in my OneNote notes. Below, I include an example of some of my initial notes taken on the interviews with researchers:

> "*To participants, recruitment of public contributors and engagement depend on a target group. I think the participants' experiences show that no one size fits all (would that be the case for all??).*
>
> *Contact with people- this is about developing long-lasting relationships with public contributors (and communities as a whole); it takes place long before the project occurs. Whose responsibility is it? Only researchers'. It seems that [this participant] thinks this is also the role of PPIE facilitators and (if not more importantly) also the university and academia as a whole. Does it map around guidance around PPI in health research?*
>
> *The power balance between researchers and public contributors. Would public contributors who my participants involve have the same perspective on what I have learnt from these interviews? Worth discussing with [my public contributors]!"*

The role of these notes was to "*critically engage*" with the data (Braun and Clarke, 2021c, p. 43).

Critically approaching the data is much easier if one is familiar with the topic or has an 'analytic sensibility' (Braun and Clarke, 2021c). This refers to the researcher having the experience to move away from describing (or summarising) the data to provide more analytic and critical thinking; for example, by linking it to the previous literature or theory. The literature on this topic was limited, but I used some theories that I discuss in a later section.[37] I attempted to keep my notes broad and record any interesting analytical aspects. There is the danger of 'drowning in' data (Seers, 2012), and I had that feeling when analysing interviews with Polish and South Asian participants – as there were forty transcripts and a wealth of data.

---

[37] See at page 137.

The familiarisation process was time-consuming, but when completed throughout the whole dataset, it created a solid analytic base on which to move to the next stages. There is no clear point when the researcher should move to the coding stage (Braun and Clarke, 2021c). It is about familiarity with the dataset and noticing nuanced patterns and relationships in data. In this thesis, this was after the second round of reading the interviews and the third reading of Twitter data. To establish more confidence in my initial thoughts, I discussed some of my notes from this stage during supervisions before I moved to the next stage, which was focused on generating codes.

7.3.2. Generating codes

Coding is a systematic process that involves reading all datasets and identifying sections relevant to the research questions (Braun and Clarke, 2006). Systematic coding means that a researcher identifies a data segment worth coding and encodes it consistently throughout the dataset (Boyatzis, 1998). Not all data sections must be coded as these might not necessarily be relevant to the study. On the other hand, some sections can have multiple codes. The development of codes is an iterative process. A researcher should track the development of codes (Byrne, 2022). This will allow a researcher to return to previous codes if they feel stuck in the analysis at any point. I made daily copies of NVivo files to record my analysis. This allowed me to come back to it and reflect on the progress and iterative process (e.g. how I refined the codes' names or how codes were grouped together). Braun and Clarke (2021c, p. 54) warn against a "*'quick and dirty' analysis*" as the fast speed and selective choice of coding can lead to not fully analysing the dataset.

Coding can be performed in Microsoft Word (Byrne, 2022), working with hard-paper copies or using specially designed software (Braun and Clarke, 2006). My initial coding was in Word using comments boxes as this was conducted with public contributors.[38] Then I imported all data to NVivo 12. The software facilitates the analysis (NVivo, 2023). It does not conduct the analysis for a researcher (Patton, 2014). Researchers have previously found NVivo to be a helpful tool to organise and assist with analysing qualitative data (Welsh, 2002). It can speed up the analysis and bring researchers closer to data (Sotiriadou, Brouwers and Le, 2014). For example, I used coding options to mark in various colours data segments for coding, wrote memos (such as notes associated with codes) and created thematic maps.

Coding can be inductive or deductive (Braun and Clarke, 2006). Inductive (also known as bottom-up) is where codes come from the data. It aims to create new ideas or explanations based on the analysed data (Patton, 2014). On the other hand, deductive coding (also known

---

[38] For more on involving public contributors in the analysis see chapter on public involvement at page 147.

as top-down or theoretical) is where a researcher codes the data based on previously developed codes, on the research area (like previous literature) or relevant theories. Deductive coding can assist with checking how appropriate current theories are to the data (Patton, 2014). My coding was inductive as there was limited understanding of my research area. However, I somewhat agree with Braun and Clarke (2021c, p. 56) that

> "*subjective and embedded process make pure induction impossible: we bring with us all sorts of perspectives, theoretical and otherwise, to our meaning-making, so our engagement with the data is never purely inductive.*"

Exploring without prior preconceptions is not only impossible but also might have a negative impact on the analysis. A researcher needs to have tacit knowledge of the research area; otherwise, it might be challenging to identify patterns (Boyatzis, 1998). I reflected on my background in the methodology chapter[39] and recognised any potential biases.

Table 4 shows an example of an initial coding extract from interviews with seldom-heard communities. The danger exists that there will be too many codes (Bryman, 2016). This means that coding will be "*too fine-grained*", and that had to be broadened to encompass more related codes (Braun and Clarke, 2021c, p. 69). Some of these codes were later linked together as these represented related issues. For example, "being the first generation in the UK" and "being a parent as time/availability constraint". Often parents can rely on their family to assist with childcare; this support network was limited (or did not even exist) for these participants who were first-generation migrants. NVivo allowed me to link these codes and record relationships which I expanded on later in my analysis. The coding process took me through multiple re-coding of the same data to ensure rigour (Braun and Clarke, 2021c).

Interview extract from South Asian participant	Examples of initial codes
"Engage that ethnic minority? So well, the first thing we need to think about how culturally we are actually directed. And if I talked about like that my community, they have come here as a first generation with so much of burden and burden in a sense that we have to complete our education. We have to get a job and it is a competition for any person in this country to get a job. So once we get a job and then it's again a ladder of the professional carrier how quickly or how much money we get through the job because of the purpose that we need. But how many	Culture impacts engagement strategies. Being the first generation in the UK

---

[39] See at page 116.

people actually do the things they love? It is all behind, is the money, so if I talk that first we struggle to come here and settle and once we settle and we are actually… we are... we are competing to get our job. (…) Well untold story that every place is still there is so much talk going on about discrimination, racial discrimination or some other things but it's still it is not on paper but we have to every time we have to actually compete in a way that say from the white British background person and I'm not talking about like European. I'm talking about the real British who are born here so we have to have like 1 1/2 times their intelligence or their talent and then only we get the rights they get. So after all these things and once we come back home from our job, we are really tired. And along with that, we have some other obligations, like if a person comes from a Muslim background and then they have a religious obligation so they have to really send their children in the evening or during the weekend to religious education. So parents are absolutely tired so then after doing all these things, they can't really focus on anything else." (Tahim, South Asian)	Finding employment is a priority. Equality not put into practice. Migrants compete with the indigenous population but no other migrants. Cultural/religious obligations. Busy lifestyle limits time for engagement Being a parent as a time/availability constraint

Table 4 Example of coding interview with a South Asian participant

Knowing when to finish coding can be challenging as there is no test to check if this stage has been completed (Braun and Clarke, 2021c). I looked at the dataset and reflected if the codes represented diverse and analytical building blocks to use as a base to construct themes. When I thought that these met this requirement, I moved on to the next stage.

7.3.3. Constructing themes,

After the whole dataset has been coded, there is a shift from looking at individual data (e.g. each interview) to identifying shared meaning across all data. A researcher looks for themes; groupings of codes that share a central organising concept (Braun and Clarke, 2021c). Generating themes is like searching for patterns in the data (Patton, 2014). Themes do not '*emerge*' from the data (Braun *et al.*, 2019). A researcher does not dig and finds a theme (that is waiting for them there to be found in the data; it is rather a process of construction (or conceptualisation) by the researcher (Braun, Clarke and Hayfield, 2022). This stage is the first step of creating themes, so one should not get too attached to these initial themes as they can later evolve or even be completely removed (Braun and Clarke, 2019). There can be multiple thematic levels (Clarke, Braun and Hayfield, 2015). The overarching theme can be to emphasise shared concepts around multiple themes. Then, the themes themselves can be divided into subthemes that share a clear relationship within one theme. Although the overarching theme or subthemes are not compulsory and should be used with caution; for example, subthemes could provide additional insight within the underlying theme (Braun and Clarke, 2023). The presentation of findings should not be fragmented as too many themes could be an example of poorly conducted analysis (Braun and Clarke, 2023).

Themes should not be confused with topic summaries. Topic summaries often appear when themes are constructed before any analysis (or even coding) has taken place (Braun and Clarke, 2021b). They might include everything said by participants on one topic but do not offer any critical or interpretative insights into the data (Braun and Clarke, 2021c). Consequently, they can be descriptive. Researchers should avoid them as they are another example of poorly conducted reflexive thematic analysis.

Thematic maps (also known as mind maps) could also enable the process of constructing themes (Braun and Clarke, 2006; Braun and Clarke, 2021c; Clarke, Braun and Hayfield, 2015). A thematic map allows a researcher to think outside the linear process and recognises the phenomenon's complexity and relationships between different ideas. This facilitates critical thinking processes (Mueller *et al.*, 2002). A researcher can use NVivo to create mind maps and link them with codes and interviews (Mortelmans, 2019). I found this option beneficial to illustrate some of my initial thoughts and relationships between themes and subthemes. However, I later used PowerPoint to produce shareable drafts with my supervisors. Figure 5 presents the initial mind map for the study using Twitter data.

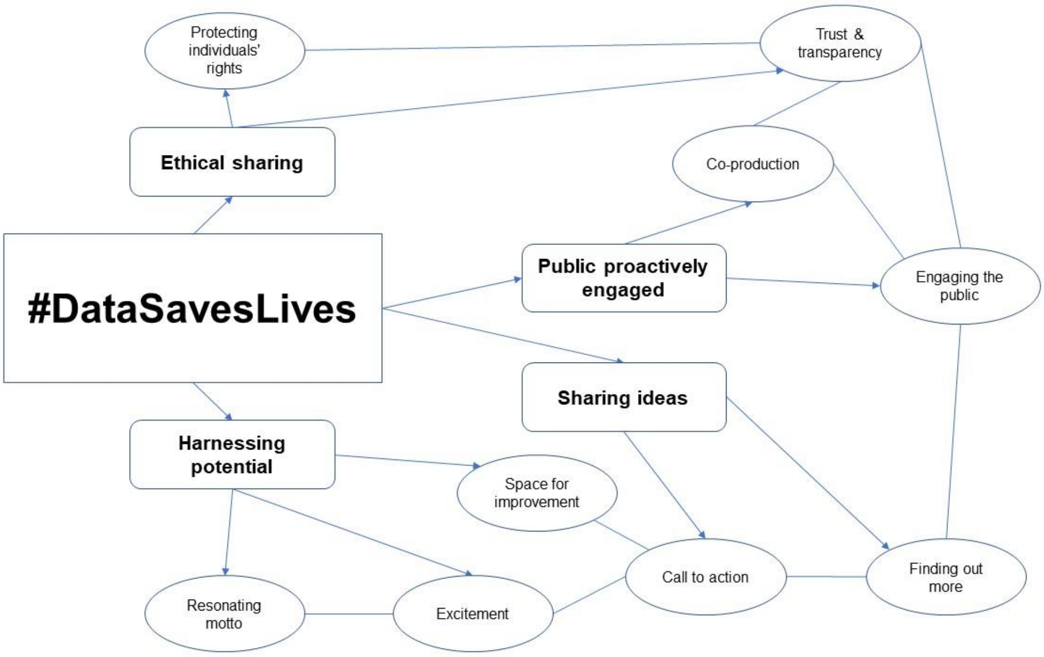


Figure 5 Initial thematic mind map for the Twitter study

From the beginning, I felt that some of these themes had too many subthemes, such as ‘harnessing the potential’ and required revisions. Initial thematic maps can develop into more refined versions (Clarke, Braun and Hayfield, 2015). This is discussed later in the next subsection.

At the end of this stage of analysis, a researcher should complete identifying candidate themes (Braun and Clarke, 2021c). These are an initial organisation of codes into a possible theme. This is only the start of the process and will require further probing as the analysis progress into the next stages of reflexive thematic analysis.

7.3.4. Reviewing potential themes,

As initial themes are generated, the next step is to come back to coded data and the whole dataset to review how these work with the candidate themes (Braun and Clarke, 2006). On the coded data level, it means re-reading all segments included under each theme (and subtheme if relevant) and reflecting on where these fit. In other words, if they represent a coherent pattern with a clear underpinning story. The theme’s concept or key message should not be based on one coded segment or one participant. When this is completed, one has to reflect if these candidate themes fit within the whole dataset. This means checking if they actually represent what is in the dataset. The thematic mapping can support the development of that process (Braun and Clarke, 2021c). Figure 6 shows how my thematic map from the Twitter study evolved at this stage.

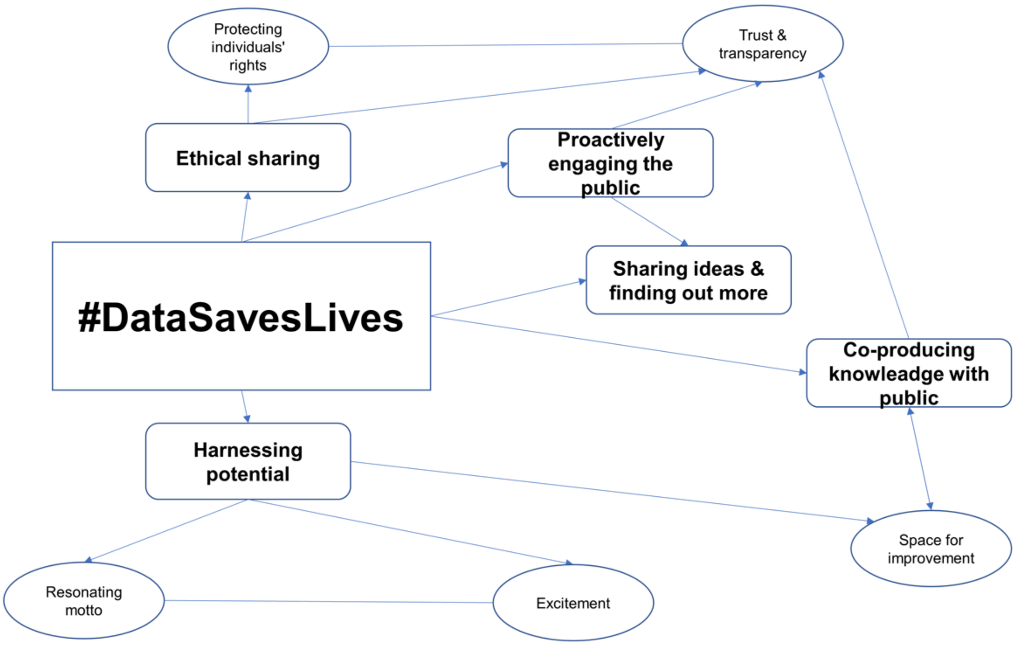


Figure 6 Revised thematic map for the Twitter study

At this stage, some themes are expected to be refined, merged, or even removed from the analysis. I added a new theme called 'co-producing knowledge with the public' that is based on the subtheme co-production (based earlier within the theme 'proactively engaging the public'). The reason behind this refinement was that after reading all coded extracts within the 'proactively engaging the public' theme, I felt that there was more than one pattern there. One focused on engagement, and the other on co-production; these are different concepts within PPI literature[40] and hence deserved to be separate themes.

At this stage, it can happen that a researcher decides that they have to take a step back in their analysis (Braun and Clarke, 2021c). My experience was similar to that of Byrne (2022), who recognised that some things said by participants offered more nuance and had to be split into additional codes. This required returning to stage two and doing some recoding. I returned to the dataset and revised codes around engagement, involvement, and co-production to recognise fine differences between these activities.

Braun and Clarke (2021c, pp. 98-99) recommend researchers consider the following questions at this stage:

- *"Is this pattern a viable theme?*

---

[40] See definitions of public involvement and engagement at page 23.

- *Can I identify boundaries of this theme? Am I clear about what it includes and excludes?*
- *Are there enough (meaningful) data to evidence this theme? Are there multiple articulations around the core idea, and are they nuanced, complex and diverse? Does the theme feel rich?*
- *Are the data contained within each theme too diverse and wide-ranging? Does the theme lack coherence?*
- *Does the theme convey something important?"*

Answering and reflecting on these questions assisted me in deciding if my themes were robust enough to move to the next stage to define and name them.

7.3.5. Defining and naming themes,

When there exists a developed thematic map, one can begin to define and name themes (Braun and Clarke, 2006). Defining is about clarifying the theme's main story, organising coded segments within that theme and creating a clear narrative. The narrative does not equal paraphrasing what data says but rather showing the nuance within it (and why this is of interest) (Braun and Clarke, 2006). It should be easy to summarise well-defined themes in a couple of sentences; in other words, it is like an abstract (Braun and Clarke, 2006; Braun and Clarke, 2021c). As this stage involves writing the narrative, it can blend with the next stage of reflexive thematic analysis that focuses on producing the report (Braun and Clarke, 2021c). My narrative descriptions of themes were used as a working base when reporting the findings to the external audience. Table 5 consists of an extract presenting my initial draft narrative of one of the themes from the study exploring how to involve seldom-heard communities in big data research.

Theme: Public being curious
The theme explores the notion that participants might not be experts in big data research, but they are curious to learn more about it. However, this does not mean that everyone will be interested in getting involved, so researchers need to reach the right people and listen more to the public on what they expect. This theme is discussed in three subthemes.
Subtheme: Patients, not experts
This subtheme elicits participants' feelings that they are not experts in this topic, so they are unsure if they should or could contribute as public advisors. However, when often discussing with them in detail what is expected of that role, they wanted to learn more, and the majority of them asked to receive details after the interview on

how to become a public advisor. Some felt that around these topics, researchers know better. *"Yeah, I think yeah it circles back to that one point, say, as long as those people are of that particular field, it would be really helpful to get their advice. Say, for example, say, you are doing some research in urology and there is a renowned doctor or staff like that in urology. And if you ask for his advice, yeah, that's good. That's good advice which will benefit both the research group and its processes and policies. But if you ask me. Me being an engineer having no concept of anatomy or anything else.*" (Onkar, South Asian)

Table 5 Example of theme narrative undertaken during reflexive thematic analysis

During the joint meeting with supervisors and public contributors[41], these initial summaries were shared with them. We discussed the analysis and then refined the themes further or linked them to the current literature. For example, we started to question if participants' perspective on members of the public not being experts by experience is not contradictory to what PPI is about. This led to further theme refinement and a nuanced conclusion.

Some initial themes' names were already in place, but this is time to reflect on if these are the final names. Themes' names should be catchy but also capture the theme's underlying message (Braun and Clarke, 2006). Although I was not so adventurous in my choices ,these can even be creative or inspired by popular culture, for example, songs or movies (Clarke, Braun and Hayfield, 2015).

The researcher can produce the report when each theme (and, if needed, subtheme) has a 'final' name and narrative abstract. I emphasise the word 'final' as reflexive thematic analysis is an ongoing process, and some changes are possible even at the next stage.

7.3.6. Producing the report

The last stage is about writing the report and sharing the findings with external audiences. This can be a thesis chapter, academic paper, or report (Braun and Clarke, 2006). The presentation of themes moves beyond describing what participants said and develops towards a critical take on what data say around the research questions. The analysis continues, and significant changes can take place here (Braun and Clarke, 2021c).

Analysis can be presented in two ways: findings and discussion in one section or two separate ones (Braun and Clarke, 2021c). The former can suit a more qualitative way of presenting data. The latter is more often associated with quantitative research. Braun and Clarke (2021c)

---

[41] See more at page 147.

recognise that journals often expect these two sections to be separate. This was also the case in my papers. However, there is a challenge that there will be a lot of repetition, so I followed their advice to include some interpretative writing in the findings section and develop on it in the discussion.

In some datasets, it might be easier to use quotes from a limited number of participants, for example, because of the way they express themselves. A researcher should avoid that temptation and ensure equal spread from the full dataset. I always ensured that illustrative quotes came from different participants. This was achieved by assigning each participant with their (anonymised) name and reporting them alongside the quotes.

The final report included extensive feedback from my supervisors and public contributors. Only when the final version is ready, the analysis in the reflective thematic analysis stops (but never fully finishes) (Braun, Clarke and Hayfield, 2022).

7.4. Orienting to data

Reflexive thematic analysis is not conducted without theoretical underpinning but rather offers theoretical flexibility (Braun and Clarke, 2021c). Despite that, conducting analysis without theoretical lenses remains one of the main misconceptions and mistakes in papers reporting that they used thematic analysis (Braun and Clarke, 2021b). Reflexive thematic analysis cannot be meaningfully conducted as an a-theoretical approach as it would lead to descriptive and unrefined findings (Braun and Clarke, 2021c). There needs to be a match between the theory and analytic claims (Braun and Clarke, 2006). In other words, when the researcher claims to be working within social constructionism, they cannot claim that the participants' experiences objectively represent the world.

There are different ways to understand theory in research (Bryman, 2016). I follow the definitions developed by Braun and Clarke (2021c). They recognise ‘big’ and ‘smaller’ theories. The big theory is about the philosophical foundation of the research. This includes how the researcher perceives reality and what meaningful knowledge is. In the case of this thesis, this was social constructionism. The second category of theories they suggest is smaller theories. These aim to explain the specific phenomenon and offer more depth to the analysis. In this thesis, two small theories were applied: models of public engagement and the Capability-Opportunity-Motivation- Behaviour model. I now discuss the role of these types of theories in my data analysis, firstly focusing on big and then on smaller theories.

7.4.1. Big theory

Braun and Clarke (2021c) see (social) constructionism as a compatible theoretical underpinning (big theory) with reflexive thematic analysis. However, they acknowledge that there is more than one approach to social constructionism, but they recognise and refer to (as do I) Gergen’s (2015) works. Gergen (2015) argues that the world is a constant flux and constructed culturally and historically. Thus, when I apply social constructionism lenses, I perceive participants' perspectives and their reality view as one of many (not as definitive views). What it means in practice is shown, for example, by Braun (2008) in her study exploring people’s perspectives on sexual health statistics around sexually transmitted infections in New Zealand. She acknowledged that data coming from participants represented their own reality, but all of these participants’ constructs need to be considered if health promotion activities can succeed.

This approach similarly applies to this thesis. In the study exploring researchers' experience of involving and engaging seldom-heard communities, I recognised that participants' accounts represented multiple realities that can be subjective (as influenced by their experience, culture, and other backgrounds). Understanding the process of how researchers involve and engage seldom-heard communities (rather than structures) would lead to a better comprehension of how different seldom-heard communities could be involved and engaged around big data research. Simultaneously, I recognised that these are not all possible realities. Other researchers (not included in the study) could have different perspectives (or seldom-heard communities with which my participants in that study were involved and engaged could have seen things differently).

The choice of big theory influences which small theories can be used in the study. They need to be complementary. Namely, small theories need to be applicable to the philosophical assumptions of big theory (in the case of this thesis, social constructionism). This is discussed in the next subsection.

7.4.2. Small theories

Critics of social constructionism would question how one could use small theories if there exist multiple perspectives and there is no such thing as the truth. However, social constructionism does not exclude small theories. Their role is not to tell the truth about the world; instead, they are useful tools to understand how the world around us is constructed (Romaioli and McNamee, 2021). This construction can lead to finding new solutions, developing practical policies and changing practices. In this thesis, I used two small theories. In the Twitter study, I applied models of public engagement and, in the study exploring seldom-heard communities' involvement in big data research, the Capability-Opportunity-Motivation-Behaviour (COM-B) model (Michie, Van Stralen and West, 2011).[42]

In the previous section, I explained that the datasets were coded inductively. The small theories were introduced to the analysis as the themes were generated. In other words, I linked the findings to the existing theories to further analytic insights. Patton (2014) argues that it is possible to conduct an inductive analysis, identify themes and then conduct an additional deductive analysis that tests how data applies to the theory. This approach of mapping qualitative themes has been successfully used in previous research. For example, Grant *et al.* (2023) mapped results from thematic analysis (following Braun and Clarke's guidance) under the constructs of normalisation process theory.

In my thesis, this process was facilitated by identifying each characteristic of the models (for models of public engagement) or constructs (for the COM-B model). Table 6 presents questions designed for each COM-B construct in relation to research questions (how to involve more Polish and South Asian communities in big data research). As the findings of the reflexive thematic analysis were mapped under the constructs of the COM-B model, it helped me to reflect on what could be put in place to improve the involvement of participants in big data research.

Constructs of the COM-B model	Capability	Opportunity	Motivation
Questions around themes to consider under the constructs.	What kind of knowledge, skills and abilities are required to become and be involved as public contributors?	What makes public involvement possible to take place? What would be the physical opportunity (e.g. appropriate resources available, accessible location or time) for participants to get involved? What social opportunities (culture and social norms) should exist to facilitate public involvement?	What would motivate participants to become public contributors?

Table 6 Questions guiding mapping of findings from thematic analysis under the COM-B model

---

[42] For more on these small theories see page 158 for models of public engagment and page 196 for the COM-B model.

This section has shown how I used big and small theories in reflexive thematic analysis. They improved the quality of the analysis by offering a theoretical lens. As this theoretical flexibility was shown as the strength of the reflexive thematic analysis, it would be appropriate to discuss in the next section the potential limitations of this analytic method.

7.5. Limitations of and misconceptions around this approach to data analysis

The reflexive thematic analysis has undertaken significant developments since Braun and Clark's (2006, p. 77) paper, in which they described it as a "*poorly demarcated and rarely acknowledged*" analytic method. Braun and Clarke themselves admit that the initial version of their method needed further work (Braun, Clarke and Hayfield, 2022). Almost two decades later, it has become a well-established and well-cited method (Braun and Clarke, 2021c). Despite the clear guidance, there remains mixed practice around applying thematic analysis. For example, Braun and Clarke (2023, p. 8) identified ten areas of problematic practice in their review of one hundred papers reporting reflexive thematic analysis in psychology journals. These include:

1. "*Undifferentiated thematic analysis.*
2. *Mischaracterising thematic analysis*
3. *Inadequate rationalisation for the use of thematic analysis*
4. *Failing to theoretically locate thematic analysis or swimming unknowingly in the waters of positivism*
5. *Misadventures with reflexivity*
6. *Inadequate description around analytic approach and process*
7. *Confusion around a deductive orientation*
8. *Conceptual incoherence around themes*
9. *Too many themes? Thinness, fragmentation and missed opportunities*
10. *Deploying theoretically incoherent quality standards*"

Inadequate use of the reflexive thematic analysis is one of the main problems in published work claiming to follow the reflexive thematic analysis. There are steps that a researcher can undertake to provide a high-quality analysis (Braun and Clarke, 2021c). First, it can be assured by a rigorous analytic process. For example, it could involve deep engagement with data at each stage of the analysis (such as I completed at the familiarisation and coding stage). Second, it requires a critical take on what data tell us. In other words, analysis is not about paraphrasing what participants said but interpreting it. This should be supported by the appropriate use of theories (as presented in the previous section). Last, a researcher should reflect on their positionality in relation to data and clearly report it in their work (which I discussed in the methodology chapter[43]).

7.6. Reflections

Qualitative analysis can take longer than expected (Braun and Clarke, 2021c). This was also my experience. The familiarisation stage felt especially time-consuming, and I wanted to dive into the data and start coding as soon as possible. However, I followed Braun and Clarke's (2006) advice of not skipping this phase. This was particularly beneficial in later stages when I referred back to my notes and reflected on them during the coding process and defining themes. I returned to my (developed) notes even when writing the thesis or papers and kept reflecting on my analysis. This is a similar experience to Trainor and Bundon (2021), who recognised that coming back to data after some time can make researchers notice new things (e.g. even returning to it as one receives feedback from peer reviewers).

Despite following Braun and Clarke's guidance on reflexive thematic analysis, I used the concept of data saturation[44] they do not recommend (Braun and Clarke, 2021d). Not following Braun and Clark's guidance does not mean that the analysis has not been conducted properly. Brooks *et al.* (2015), who offer another thematic analysis (called template analysis), feel "*that it is crucial that researchers are not precious about "their" ways of working with thematic analysis*". Braun *et al.* (2019, p. 849) "*somewhat agree"* with that statement. In the conversation on reflexive thematic analysis, Clarke says that they provide "'*a walking stick, a pair of socks, and a compass; plot your own route*'" rather than a map (Braun, Clarke and Hayfield, 2022, p. 431). Brooks *et al.* (2015) go even further and argue that guidance (e.g. by Braun and Clarke) can be particularly helpful for novice researchers, while more experienced academics could draw on different approaches. This could encourage mixing different methods. For example, Watson and Douglas (2012) used aspects of thematic analysis and grounded theory in their analysis of how neighbourhood impacted mental wellbeing among disadvantaged young people. Consequently, a researcher could use a new (modified) analytic method that might be appropriate for the particular study. Bryman (2016), in his textbook on social research methods, presents the process of thematic analysis based on principles defined by Braun and Clarke but also brings insights from other researchers. However, except for data saturation, following Braun and Clarke's reflexive thematic analysis allowed me to keep a clear focus and (flexible) structure on how I analysed data.

---

[43] See my reflection on my positionality at page 116.

[44] See the discussion on data saturation at page 114.

7.7. Conclusions

This chapter presented the process of analysing qualitative data. I justified why reflexive thematic analysis was the most appropriate analytic method by especially emphasising its theoretical flexibility. I discussed each step of the analysis, reflecting on how I applied them in practice. Further, I explained my orientation to the analysis. Then, I considered the potential limitations of this analytic approach. The chapter concluded with some reflections on the analysis and using reflexive thematic analysis in my research.

8. Public involvement in the thesis

8.1. Introduction

Involving public contributors has become a more widespread practice among researchers, including doctoral students (Manikandan *et al.*, 2022; Jones and Hunt, 2022; Foley *et al.*, 2021; Dawson *et al.*, 2020; Tomlinson *et al.*, 2019; Troya *et al.*, 2019). As discussed in the literature review, public contributors' involvement has the potential to shape the research project significantly.[45] Public involvement in my thesis aimed to ensure that there was a lay perspective and the research findings were relevant of to the general public, especially Polish and South Asian communities.

This chapter discusses how I incorporated public involvement into my research. First, I discuss the recruitment and the support public contributors received. Then, I reflect on their involvement in each study. This is followed by a broader discussion around the challenges of public involvement during doctoral research. Last, I deliberate on how the Covid-19 pandemic changed the way I worked with public contributors.

8.2. Setting up and recruitment through the Public Advisors Forum

Prior experience working with public contributors is not needed for doctoral students to ensure meaningful involvement (Rees, Sherwood and Shields, 2022). Previously, I worked in a charity where we held public and patient engagement sessions, so I had some understanding of PPIE. This was further developed through training sessions and webinars I attended. My supervisors had prior experience involving public contributors in research. I received further assistance from the PPI manager with NIHR ARC NWC, who is a conduit between public contributors and researchers. No ethical approval for public involvement was required. However, I recognise that this might be a challenge at institutions that require it (see, for example, the experience by Rees, Sherwood and Shields (2022)).

Three months through my PhD journey, I recruited public contributors through the NIHR ARC NWC Public Advisors Forum.[46] As my PhD research did not require a specific experience as a patient of a particular health service, it was an appropriate place to find public contributors. NIHR ARC NWC offers inductions (on being a public contributor), regular support, training and networking opportunities. Public contributors attend the Public Advisor Forum, regular local conferences (called ARCFests), joint doctoral students and public contributors journal club and are invited for relevant training opportunities. Most of these activities are reimbursed following NIHR rates, and any additional expenses are covered (e.g. travel costs if the meeting

---

[45] See more at page 34.

[46] Within ARC NWC public contributors are referred to as public advisors. However, for consistency in the thesis I keep using public contributors.

takes place face-to-face and printing). There is an established process to be reimbursed for their time, which the ARC administrator supports. This well-established network[47] of support ensured that I did not have to provide any training or explanation of what being a public contributor entailed but focused on their involvement in my PhD research.

After an initial chat with the PPI manager, I decided to open the opportunity to all public contributors registered with ARC NWC. Through the mailing list, I sent a role descriptor providing a short overview of the project, what involvement would entail and reimbursement (see Appendix 7 for the full email). Interested public contributors were asked to email me a short paragraph explaining their interests. As more than two people volunteered, I chose candidates who considered themselves a part of a seldom-heard community, as that lived experience was the most appropriate for this research project. Consequently, two public contributors from South Asian backgrounds joined me on my doctoral journey.

I initiated the involvement process, but it was a joint decision with public contributors on how much they were involved in the different research projects. The reason for letting public contributors choose was based on the premise that involvement should be an interesting opportunity. For example, public contributors could be interested in learning more about the research (as methods or topics) (Newington *et al.*, 2022). After discussion of different options, public contributors have been involved in the two-stage review, interviews with seldom-heard communities and interviews with researchers. Due to funding restraints, it was not possible to involve them in the Twitter study. I first discuss their involvement in the reviews, followed by the qualitative studies. However, these activities were often happening simultaneously. Table 7 summarises the impact of involving public contributors in these research projects. This has been guided by the GRIPP2 reporting form to ensure transparency and quality (Staniszewska *et al.*, 2017b).

Project	Research activity	Public contributors' involvement	Training offered	Impact of involvement
Review	Design	Involved in designing review questions. Co-designing the system logic model underpinning the review. Reviewing and commenting on review protocol.	Training around big data research and reviews. Additional reading on reviews was offered for one reviewer.	Clarified the focus of the review and inclusion criteria. Additional search terms.

---

[47] To read more about its background and current strategy, please see https://arc-nwc.nihr.ac.uk/wp-content/uploads/2021/03/ARCNWCPPCIEPSStrategy.pdf

		Co-authorship of review protocol.		
	Screening stage	Involved in jointly screening a sample of titles, abstracts and full papers.	Training before each screening stage: title, abstract, full paper.	Ensured appropriate inclusion of papers in relation to the review aims.
	Data extraction	Piloting and improving the data extraction form. Involvement in checking the extracted papers.	Training on data extraction and jointly extracting data from a sample of papers.	
	Analysis	Feedback on findings.	None.	Ensuring relevance to members of the public and how findings relate to their experience as public contributors.
	Dissemination	Co-authorship of the conference abstract.	None.	Ensured public views in the output.
Qualitative studies	Design & ethics application	Involved in the ethics application. Reviewing participant information sheet, consent form and interview guide. Co-designing the study ads. Co-design and co-chair PPI sessions with lay public members to pilot interview questions. Recruited participants for the South Asian group.	None.	Shaping the study aims. Ensuring lay language (no jargon) in all study materials. Creating appealing study ads to reach potential participants. Changes to the interview guide.
	Recruitment	Acted as gatekeepers to assist with the recruitment of South Asian interviewees.	None.	Assisted with reaching South Asian participants. Recruitment of participants for each group (Polish and South Asian participants'

				recruitment targets were met).
	Analysis	Coding of one interview per participant group. Involved in shaping themes (e.g. through joint meetings with the supervisory team).	Training around the reflexive thematic analysis and being reflective.	Involvement shaped the findings and the priorities for the discussion section.
	Dissemination	Reviewing and commenting on papers. Co-authorship of papers. Co-authorship of conference abstract and co-presenting it. Advised on disseminating study findings to the public.	Additional reading on research methods was identified for one public contributor.	Assisted in identifying the target audience for dissemination. Ensured public views in all outputs.

Table 7 Summary of the public involvement, training offered and impact in the thesis.

8.3. The literature review

This was the first project public contributors became involved in within my PhD thesis. Public contributors might require support and training to be fully involved in research (Cowley *et al.*, 2019; Hemming *et al.*, 2021). Big data research can be a complex issue, so during the first meeting, I delivered training on this research field to ensure their familiarity with it. This made the public contributors feel confident in discussing this topic with me. We also used the session as an opportunity to build the public involvement and engagement system logic model[48] in big data research.

We jointly discussed if they would be interested in getting involved in both stages of the review (scoping and systematic). The answer was positive, and they became co-authors of the review protocol[49]. As co-authors, the public contributors provided a lay perspective and thus shifted the focus of the review to offer more practical implications for members of the public.

Then, we agreed that they would contribute at each review stage. I provided additional training sessions before title, abstract and full paper screening, and data extraction. These sessions focused on both theoretical and practical research skills. For example, I spoke about the process during the title screening stage. Then, as a group, we jointly screened one hundred titles and discussed why we wanted to include them in the abstract stage. Afterwards, I sent everyone their titles for screening within their own time. Each person completed the screening, and I compared them with one another to identify disagreements. Thereafter, we reconvened

---

[48] See the initial system logic model at page 63.

[49] See chapter 4 at page 59 for a copy of the paper.

as a group and discussed these disagreements before moving to the next stage. The process was repeated at the abstract and full paper stages. Most disagreements happened at the title and abstract stages, as it was often unclear if the paper had any PPIE aspect based on the information provided. After discussion, we jointly reached a consensus to include these screened papers for full paper review. I organised titles and abstracts for screening in Word documents and did not use any screening software to avoid introducing new programmes to the public contributors.

After the data extraction stage, I gathered the findings and organised a meeting with public contributors and another doctoral student involved. During the meeting, I shared some findings, and after each section, we jointly discussed their relevance to the review aim and their implications for both researchers and the members of the public. Again, public contributors offered a public perspective. For example, during the meeting, we discussed that only a few included papers cited public involvement guidance.[50] They felt that this is not a surprising but disappointing finding. However, after reflection, we agreed that this could be explained by the fact that the involvement and engagement guidance for big data research has been published only recently. Later, they also commented on the system logic model[51] developed through the review and said that it is representative of their experiences as public contributors. At this stage, they particularly welcomed adding the section on barriers in the model because they felt that researchers often had not discussed these when public contributors became involved in new projects.

Public contributors co-authored the conference abstract reporting some findings from the review. This was accepted at the Administrative Data Research Conference 2023.

8.4. Interviews with Polish and South Asian communities

Public contributors were involved from the design stage of this research project focusing on Polish and South Asian communities. Usually, we had joint meetings. However, due to time commitments, it is not always possible to meet all public contributors each time (Manikandan *et al.*, 2022). When that happened, I held one-to-one meetings. As qualitative studies require ethical approval, public contributors were involved in the ethics application. They reviewed interview guides and consent forms and helped design and supported the recruitment for broader PPI discussion. The ethics application was submitted at the end of the first year, so public contributors felt that they had become familiar with big data research. This made them wonder if everything in the topic guide was written in an accessible language or if they easily

---

[50] See this review findings at page 83.
[51] See the updated model at page 90.

understood it because of their familiarity with the research. On reflection, we jointly decided to run another two focus groups with the members of the public not involved in the thesis to pilot the interview questions. Public contributors were involved in co-designing them. The South Asian group was co-chaired with one of the public contributors. However, as the Polish group was conducted in the participants' mother tongue (that public contributors do not speak), I asked for help from a Polish speaker experienced in public involvement from one of the charities working with migrants and ethnic minorities in the UK. These two PPI groups took place online, but we designed them to be engaging and interactive, for example, through visual minutes, Padlet and polls (Teodorowski, Ahmed and Kasprzyk, 2021)[52]. Appendix 8 shows the visual minutes summarising the discussion. These groups ensured that all questions were in an accessible language; thus, interviewees felt comfortable answering these questions.

Again, we discussed if public contributors would like to be involved in other stages of that study. We reflected on the time commitment each task might require as public contributors asked me to estimate how long it would take, and based on that estimate, they decided how much they would like to be involved. Therefore, they got involved in coding one interview for each participant group (Polish and South Asian) and defining themes.

There is evidence in the literature that public contributors can contribute to qualitative analysis (Garfield et al., 2016). For example, Hemming et al. (2021) reflected on the process of training and involving one public contributor in thematic analysis (following Braun and Clarke's method, as myself). One of the things they noticed was that there was a significant overlap between codes identified by public contributors and the researcher. This was my experience as well.

In line with the established process, I delivered training on reflexive thematic analysis. During the session, we jointly coded an extract from one of the interviews and reflected on what and why we coded these sections in reference to the research questions. Afterwards, we separately coded one interview and came back together to discuss codes. This was also used to reflect on whether the interview guide was appropriate. Rubin and Rubin (2005) recognise that asking every follow-up question is not always possible as an interview has time constraints. The interviewer has to decide which avenues of inquiry must be followed and choose those most relevant to the research questions. Public contributors assisted in the process of identifying key issues for exploration with participants. In the coded interview, the interviewee suggested that religious events could be a good place for researchers to engage communities in big data research. Public contributors questioned this method's

---

[52] See more discussion on how successful these interactive digital tools can be during online PPI discussions at the blog post by Teodorowski, Ahmed and Kasprzyk (2021)

appropriateness and suggested exploring this issue with other participants. Thus, I asked the following interviewees who identified themselves as Muslims about their views on it. This allowed eliciting the complexity of this kind of engagement during religious festivities.

When I completed coding the remaining interviews and organised the initial themes, these were sent to the public contributors and supervisors. We decided to have a joint meeting and discuss the initial results. Based on that discussion, I drafted a conference abstract (Teodorowski *et al.*, 2022a) and a paper[53], both of which public contributors co-authored. The conference abstract was accepted, and we received additional funding for one public contributor to co-present the findings with me at the Population Data Linkage Network Conference 2022 in Edinburgh.

8.5. Interviews with Researchers

The qualitative study exploring the researchers' perspective was conducted alongside interviews with Polish and South Asian participants. Therefore, public contributors were involved in its design and in the ethics application (as it was a joint one). We followed a similar process for the analysis, and they coded one interview. Again, public contributors identified additional issues for exploration. One participant spoke about the seldom-heard community, but it was unclear how they would define their public contributors as seldom-heard. This was explored as an additional follow-up question during other interviews. As the analysis took place during the summer period, arranging a joint meeting between supervisors and public contributors was challenging, so we decided to discuss initial themes in separate groups. Thereafter, I developed the themes and wrote a paper and the public contributors were involved as co-authors.[54]

8.6. Challenges of public involvement in doctoral research

8.6.1. Power relationships

Doctoral research ownership remains challenging when it comes to public involvement (Jones and Hunt, 2022). PhD research is supposed to be a student's work, but simultaneously involvement principle requires sharing power over the research process. I aimed to make the public contributors feel like part of the research team so they felt like co-researchers and equal partners (as this is a collaborative process) (Renedo *et al.*, 2015). The public contributors had prior experience in being involved in research projects (and academic publications). This meant that they were confident in voicing their opinions and challenging mine (Newington *et al.*, 2022). Wherever possible, we made decisions by consensus. Similarly to Rees, Sherwood

---

[53] See chapter 11 at page 192 for the copy of the paper.

[54] See chapter 10 at page 174 for the copy of the paper.

and Shields (2022), I recognise that building consensus is not always possible. If I could not incorporate the public contributors' feedback, I explained why, and we found an alternative. For example, one of the public contributors commented on the draft paper reporting findings on interviews with researchers that it felt too hard for a lay audience to understand. We discussed that the target audience for that publication was researchers and agreed to produce an additional lay summary which would be accessible to the public.

My concerns were also around their understanding of the topic or research. For example, after the discussion on the methodology used in the review, I identified additional reading which helped to improve the public contributor's knowledge of the topic.

Public contributors need to be recognised for their involvement and feel valued (Newington *et al.*, 2022). Following NIHR's (2022) guidance, public contributors were reimbursed for their research contributions. However, financial incentives are often not the main driving factors for those involved in health research (Newington *et al.*, 2022). In my research, the value of being involved was not only monetary, as they had an opportunity to learn new skills, be co-authors on papers, blogs, and co-present research findings at conferences. In addition to the previously mentioned conference, the second public contributor presented with me at the online conference of the Polish Academy of Science on Ethics in Social Research on public involvement in the ethics process (that was discussed earlier in this chapter).

8.6.2. Positionality of public contributors

In the methodology chapter, I discussed reflexivity and reflected on my positionality in relation to this thesis.[55] Public contributors can be biased in their opinions (Rees, Sherwood and Shields, 2022). Limited literature discusses how public contributors' experiences influence what they perceive in data. However, some concerns exist that strong-minded individuals could attempt to bring their agenda to research (Becker, Sempik and Bryman, 2010; Garfield *et al.*, 2016). This is especially recognised in activist research, where there can be multiple identities (e.g. being a researcher and an activist) (Bloom and Sawin, 2009). Public contributors share their lived experiences and are influenced by them (similarly to researchers). That potential bias could impact the quality of research. To overcome this challenge, I provided training around reflexivity for my public contributors. It focused on the importance of reflection and how our experiences, backgrounds and personal views could affect what we perceive in data. Then we jointly discussed each of our positionalities. Public contributors asked for some questions to consider when analysing interviews. These were influenced by the work of Helen Kara (2020):

---

[55] See at page 116.

1) What views, perspectives, and opinions do I bring to the discussion?

2) Are my views similar or different to the participants' experiences?

3) Are there other perspectives to understand what participants said?

We started each analysis meeting with a short conversation about public contributors' initial thoughts on the data and their positionality towards it. This approach was similar to Hemming *et al.* (2021) who provided the public contributors with a reflective guide consisting of twenty questions as guidance. The impact of my training was visible during meetings. Public contributors started to question why they looked at the data extract in a particular way and considered their reactions towards it. For example, in relation to the previously mentioned example of religious events, they linked their personal experiences and considered the participant's answers. As the public contributors came from South Asian communities, they offered an insider view into this participant group. This was a unique perspective, as neither my supervisor team nor I could have provided those insights.

8.6.3. Researching within the confines of the field

Writing and conducting doctoral research must be within the field's style and confines. Public involvement has become established among some researchers but sometimes remains misunderstood or used in tokenistic ways. Meaningful involvement requires researchers to believe that public involvement would benefit their research and improve the chance of publications and career progression. Writing a doctoral thesis differently than the established approach might face mixed opinions among peers (Weatherall, 2019); for example, this might be the case when doctoral researchers actively involve public contributors throughout their research. If the researcher is in a supportive environment towards public involvement, there is a higher chance that public involvement will take place and be meaningful to researchers and public contributors. Boylan *et al.* (2019) interviewed researchers to understand their PPI experiences. Some participants described their colleagues' views around public involvement as cynical, sceptical or ambivalent. Thus, the organisational culture and support among senior researchers for public involvement can influence how involvement takes place and thus, its impact (Boylan *et al.*, 2019). Supportive supervisors can facilitate the public involvement process (Rees, Sherwood and Shields, 2022). In my experience, my peer network (as they also had public contributors involved in their research) and supervisors offered positive feedback on public involvement in my PhD thesis.

8.6.4. Working with public contributors as a learning process

Working on the thesis is when doctoral students develop their own unique scholarly identity (Hay and Samra-Fredericks, 2016). My perspective changed too to becoming more academic, but keeping that lay view throughout health research is important because it ensures that findings could have real-life implications and thus benefit the broader community. Public contributors ensured that I reflected on the relevance of my research to the broader public. This can be especially seen in co-authoring or co-presenting with public contributors. Peer writing with fellow doctoral students offers an opportunity to develop academic skills (especially writing and critical thinking) as they bounce ideas with similarly minded colleagues (Adamek, 2015; Wegener, Meier and Ingerslev, 2016). Working with public contributors can, and in my case, had the same impact on my development as a scholar and helped create new ideas.

8.7. Impact of the Covid-19 pandemic

The Covid-19 pandemic changed public involvement as most meetings moved online (Clark, Vliet and Collins, 2021). Online public involvement meetings have advantages and drawbacks (Manikandan *et al.*, 2022). Growing evidence shows that public involvement can be conducted remotely (Jones *et al.*, 2022a). Public contributors could join meetings from the comfort of their homes, and no travel meant they could fit the discussions around their other commitments. This could benefit public contributors that were not involved because of distance and travel time (Clark, Vliet and Collins, 2021). This was the case in my PhD, as my public contributors admitted that it would not have been possible for them to be involved in many projects if there was a need to travel between locations rather than ‘move’ between Zoom meetings. However, we also sometimes faced IT challenges (e.g. poor connection), or if using creative tools (such as whiteboard, polls, Padlet), the accessibility to these differed on the phone, tablet and computer. Public contributors could also have printed materials and be reimbursed for them if needed. Our first face-to-face meeting was one year after the initial session. We decided to go for a social gathering to get to know each other better as we knew that the remaining meetings would continue online (due to the travelling barrier). At the end of my doctoral journey, I also invited them for an in-person celebration meeting at a local restaurant to reflect informally on our joint journey.

8.8. Conclusions

This chapter has shown that the involvement of public contributors ensured that my thesis was relevant to the public. They influenced the study design (interview guide, consent forms) and the recruitment strategy and helped to provide a more nuanced analysis. The ongoing involvement also shaped my thinking as a researcher, as I learnt from the public contributors (Manikandan *et al.*, 2022). In this thesis, public contributors had an active role as co-researchers or critical friends. This was in contrast to some other involvement in doctoral research where public contributors acted only as a “*patient supervisor,*” a member of the supervisory team who guides the work (Jones and Hunt, 2022, p. 5).

9. Twitter study

This chapter presents verbatim the first primary study findings from the paper published in the Journal of Medical Internet Research titled *Use of the Hashtag# DataSavesLives on Twitter: Exploratory and Thematic Analysis* (hereafter “Twitter study”) (Teodorowski *et al.*, 2022b). Before presenting the Twitter study, I explain how this research builds upon the literature review; then, I briefly outline the aims and how findings contribute to the current knowledge.

9.1. Linking the literature review with the paper

The literature review identified the following issues regarding public engagement around big data research:

- Social media (e.g. Facebook or Twitter) could be used to share research with the public, but there is limited evidence on how to do it.
- Nation-wide campaigns could offer an opportunity to reach the public about the benefits of big data research. However, how to do it in a meaningful way remains unexplored.
- The UK experienced public controversies around the linkage of health data for research.

The Twitter study addressed these gaps in the literature. It focused on the Data Saves Lives campaign that highlights the benefits of big data research. Previous research explored the campaign (Hassan, Nenadic and Tully, 2021). However, the study coverage was between 2016- 2017. My research built on that study and explored how the campaign was utilised in 2021. The context and timing of this research are of particular importance. First, the campaign had been relaunched internationally. Second, the UK government adopted the hashtag and used it to promote their new data linkage project (General Practice Data for Planning and Research). As I discussed in chapter 3,[56] that project faced public controversy and has been postponed.

Therefore, the Twitter study aimed to understand how the campaign had been presented on Twitter as the UK government adopted it and what attitudes toward the campaign were shared among Twitter users. To better understand that, the study was underpinned by three models of public engagement (deficit, dialogue and participatory models). All results from the thematic analysis were mapped to one of these models. Further analysis included assigning each message a positive, negative or neutral attitude towards to campaign. Also, descriptive statistics were used to summarise data.

---

[56] See more at page 53.

The findings in this paper assist in understanding how the Data Saves Lives campaign was utilised to engage the public around big data research. The lessons from this engagement could also be relevant to other social media campaigns. First, the findings showed that the Twitter users' response was mostly positive towards the campaign, and the discussion remained UK based. Second, findings indicate how the campaign mostly remained within the deficit and dialogue models of public engagement. Therefore, the paper called for more activities within the participatory model.

9.2. Abstract

Background: The hashtag DataSavesLives is a public engagement campaign that highlights the benefits of big data research and aims to establish public trust for this emerging research area.

Objective: Exploring how the #DataSavesLives is utilised on Twitter. We focused on the period when UK government and its agencies adopted the #DataSavesLives in an attempt to support their plans to set up a new database holding NHS users' medical data.

Methods: Public tweets published between 19th April and 15th July 2021, using the hashtag DataSavesLives were saved using NCapture for NVivo 12. All tweets were coded twice. Firstly, each tweet was assigned a positive, neutral or negative attitude towards the campaign. Secondly, inductive thematic analysis was conducted. The results of thematic analysis were mapped under three models of public engagement: deficit, dialogue and participatory

Results: Of 1026 unique tweets available for qualitative analysis, discussion around #DataSavesLives was largely positive (n=716) or neutral (n=276) towards the campaign with limited negative attitudes (n=34). Themes derived from the #DataSavesLives debate included: ethical sharing, proactively engaging the public, co-producing knowleadge with public, harnessing potential and gaining an understanding of big data research. The Twitter discourse was largely positive towards the campaign. The hashtag is predominantly used by similarly-minded Twitter users to share information about big data projects and to spread positive messages about big data research when there are public controversies. The hashtag is generally used by organisations and people supportive of big data research. Tweet authors recognise that the public should be proactively engaged and involved in big data projects. The campaign remains UK centric. The results indicate that the communication around big data research is driven by the professional community and remains one-way as members of the public rarely use the hashtag.

Conclusions: The results demonstrate the potential of social media but draws attention to hashtag usage being generally confined to 'Twitter bubbles'; groups of similarly-minded Twitter users.

Keywords: consumer involvement; patient participation; stakeholder participation; social media

9.3. Introduction

Well-established ways for sharing knowledge with the general public by researchers include academic publications, presentations or media engagement (to name a few). However, previous research has raised concerns that the communication between scientists and the public needs to be more accessible and interactive than traditional engagement activities (Kurath and Gisler, 2009; Simis *et al.*, 2016; Priest, Goodwin and Dahlstrom, 2018). Public engagement, when it is a two-way process of sharing, promoting and disseminating research to the public (INVOLVE, 2020; National Co-ordinating Centre for Public Engagement, 2021), can improve trust between researchers and the public (Aitken, Cunningham-Burley and Pagliari, 2016). The growth of social media platforms such as Twitter, a microblogging platform (up to 280 characters per post) (Twitter, 2022) offers a more interactive way to engage with the public, and can be particularly useful in promoting engagement around controversial topics. Twitter provides a less formal and more dynamic interaction among its users. Posts (tweets) are open to read for everyone but only Twitter users can post (tweet) them (but Twitter is free and easy to sign up to). Users can reshare original tweets (retweet) with their audience (followers). Researchers are already active on Twitter to communicate their work as they can reach the public (López-Goñi and Sánchez-Angulo, 2018), colleagues in their field (Fuller and Allen, 2016), policymakers and practitioners (Tripathy *et al.*, 2017).

One of the key issues in big data research and one subject to a prolonged public debate, is the reuse of medical data for research. Often called big data, it has the potential to provide novel health solutions and improve health inequalities (Raghupathi and Raghupathi, 2014; Hemingway *et al.*, 2018). Non-(re)use of data can negatively impact health services and research (Jones *et al.*, 2017). However, some public members are concerned about how their medical data is stored, controlled, (pseudo)anonymised and reused (Aitken *et al.*, 2016; Kalkman *et al.*, 2019b). Public trust and support are needed for big data projects to continue (Carter, Laurie and Dixon-Woods, 2015). However, there remains little public understanding of big data research (Aitken *et al.*, 2016).

Data Saves Lives, is a public engagement campaign, that highlights the benefits of big data research, showing how patient data can be used securely to improve healthcare (Data Saves

Lives, 2020). The campaign tries to build trust between researchers and the public. It was started by the University of Manchester's Health eResearch Centre in 2014. Since then, it has expanded outside the UK, and in 2019 was launched in Europe. The Data Saves Lives European initiative is a multi-partner project led by the European Patients' Forum and the European Institute for Innovation through Health Data (European Patients Forum, 2019). The campaign activities target social media, especially Twitter, using the hashtag #DataSavesLives. Hashtags allow the linkage of all posts on the same subject. Any user can use hashtags on Twitter, and to gain broad coverage, it is recommended to get as many Twitter users as possible to use the hashtag. However, this also means that the hashtag's originators do not control by whom and how it is used. This can lead to high-jacking of the hashtag by other users, who may use it for different purpose than initially intended (Bradshaw, 2022; Jackson and Foucault Welles, 2015).

In 2021, the UK government and its agencies adopted the hashtag #DataSavesLives to support their plans to set up a new national database holding NHS users' medical data, which could be, in some circumstances, available for sharing with third parties (Scott, Emerson and Henderson-Reay, 2021). The idea was driven by the Covid-19 pandemic and the recognition that data has the power to shape and improve healthcare services (Department of Health & Social Care, 2022b). The plan was to collect 55 million patients' pseudonymised data in England to be reused (for example, to support services and research). This received heavy criticism from activists regarding lack of transparency around informed consent and confidentiality (O'Dowd, 2021). Patients would have only limited time to opt-out of the scheme, and their consent was mentioned only once in the initial governmental policy documents (Shaw and Erren, 2021). The plan's legality was challenged, and there were concerns that medical professionals would refuse to comply by not sharing their patients' data (Armstrong, 2021). Poor communication resulted in public concerns around this new scheme. British media outlets from the Independent to the Daily Mail, described the plan as "controversial" (Shaun, 2021; Craig, 2021). These attitudes were not new, as a similar (but not linked) project was abandoned in the past due to negative public opinion (Hays and Daker-White, 2015; O'Dowd, 2021). Medical professionals had raised concerns about building trust with the public regarding new government plans. The British Medical Association and the Royal College of General Practitioners called for a better public engagement campaign to alleviate public fears (BMA and RCGP, 2021). One and half million people initially opted out of the scheme (Wise, 2022). The government deferred the deadline for the public to opt out of the new database scheme due to public concerns (NHS Digital, 2021). Later, the policy was reviewed to discuss building trust with the public further (Goldacre and Morley, 2022). The new governmental policy was published June 2022 (Department of Health & Social Care, 2022a). In contrast, there have been no such controversies in Europe or adoption of #DataSavesLives by European public institutions.

Previous studies explored public perceptions of big data research, but few have examined how online public engagement campaigns could promote the benefits of big data research. One paper discussed #DataSavesLives on Twitter, but its coverage was between September 2016 to August 2017 (Hassan, Nenadic and Tully, 2021). Our study expands on previous research and explores how the campaign's hashtag was used when the UK government decided to adopt the hashtag in its campaign strategy. Thus, we cover the period of April to July 2021, when there was an ongoing discussion in news headlines around the newly proposed scheme.

9.4. Models of public engagement

Science communication as a research area emerges from diverse fields and offers theoretical underpinnings for how researchers can engage with the public (Priest, Goodwin and Dahlstrom, 2018); where the public is understood as any person in society (Burns, O'Connor and Stocklmayer, 2003). We will use the terms public and public members in this paper as people who do not have a background in healthcare or big data research, lay people. Three theoretical models of how researchers can engage with the public exist in the literature: deficit, dialogue and participatory (Metcalfe, 2019; Trench, 2008). These differ in where they locate researchers or the public in the process of engagement (Trench, 2008).

9.4.1. Deficit model

The deficit model is the oldest and nowadays heavily criticised model for being too passive a form of communication (Burns, O'Connor and Stocklmayer, 2003). It is also known as the knowledge transmission model (Seethaler *et al.*, 2019) as it assumes that the public has limited understanding of the research, and through engagement, researchers can educate the public and explain the complexity of their work, promoting a researcher-centred model (Simis *et al.*, 2016; Bucchi, 2008). The model theorises that if the public is not supportive of the ongoing research, researchers only need to explain it better to the public (Wynne, 2006; Bucchi, 2008). Thus the underpinning problem is public's lack of understanding (Priest, Goodwin and Dahlstrom, 2018). The weakness of this model is the ongoing need to educate the public, which can be only done through a top-down (and usually one-way) approach with researchers giving the public information and telling them how they should understand the issues. The empirical evidence has shown that the deficit model of engagement does not change public views towards science (Miller, 2001).

9.4.2. Dialogue model

The dialogue model was developed in response to the mistrust the public had in research in general (but particularly in medical research) and the perceived failure and passivity of the deficit model to tackle that challenge successfully (Wynne, 2006). Public and researchers may have different perspectives and can interpret the same things differently (Bucchi, 2008). The dialogue model recognises a need for an active exchange between researchers and the public, ensuring two-way communication (Trench, 2008). This communication can improve understanding among both groups as they can see different perspectives on the same issue. The dialogue model moves away from researcher-centredness in the communication process and invites public views on the research. Public understanding of science is no longer perceived as limited or inferior to researchers' (as it was in the deficit model) but rather it is perceived to offer a unique view. The model theorises that the dialogue can further improve trust if researchers listen and implement public feedback. The public will not only understand the researchers' perspectives better but also will be more willing to act upon on their advice (Lee and Garvin, 2003). For example, they might be more willing to take a new medicine or participate in research.

9.4.3. Participatory model

Shifting further the power balance between researcher and public, the participatory model argues for public-centredness in communication. Researchers and the public discuss the research agenda, and in contrast to the dialogue model, they also jointly find solutions. This democratisation of the process has been argued to have the potential to improve the quality of information and reaching the public (Carney *et al.*, 2006). Both groups have something to gain from this cooperation (Trench, 2008). In health research, it would come under the definition of public involvement, where work is being done together *with* the public rather than *for* them (NIHR, 2021a). Growing research shows that public contributors (e.g. lay members) are successfully involved in developing and shaping engagement of healthcare services (Mockford *et al.*, 2012).

9.5. Research questions

Underpinned by the (deficit, dialogue and participatory) models of public engagement, the current study aims to answer the following research questions:

1. How was the #DataSavesLives used on Twitter as the UK government adapted the hastag in its campaign strategy?
2. What were the attitudes towards the campaign among Twitter users utilising #DataSavesLives?

9.6. Methods

9.6.1. Data collection

Tweets were recorded using NCapture software for Google Chrome. This web browser extension collects social media data such as tweets (including retweets) and imports them directly to NVivo 12 for analysis. Only public tweets from the previous week could be recorded. NCapture does not guarantee that all tweets can be captured at once as this depends on Twitter; thus, we captured tweets twice per week (Tuesday and Thursday) to get maximum coverage. If an individual tweet is captured twice, NVivo 12 uploads it into the dataset only once. Tweets using the hashtag #DataSavesLives were captured for three months between 27th April to 15th July 2021. This covered tweets that were posted between 19th April and 15th July 2021. A total of 3638 tweets (including retweets) were collected. We cleaned the dataset in the NVivo 12 (see Figure 7). All retweets, duplicates, tweets consisting only of hastag, spam, tweets in other languages than English were removed. After cleaning the dataset there 1026 tweets were used in the qualitative analysis.

Data saturation was deemed to have been reached. This assumption is based on previous research, which successfully conducted a qualitative analysis of fewer than 1000 tweets and provided novel insights into the online discussion through Twitter hashtags (Berry *et al.*, 2017; Bogen *et al.*, 2022; Rashid, Mckechnie and Gill, 2018).

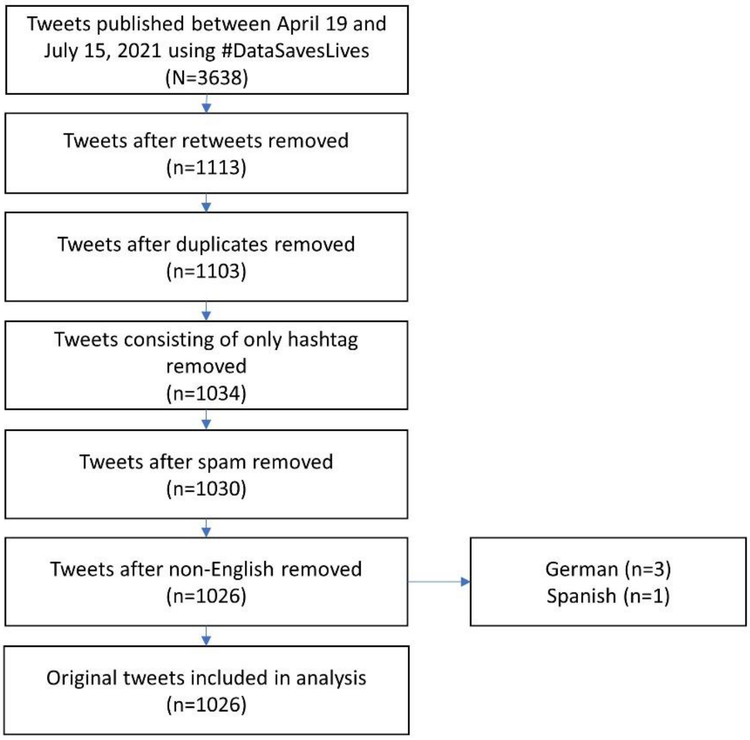


Figure 7 Process of cleaning data sets for qualitative analysis

9.6.2. Analysis

The analysis was conducted using NVivo 12, NCapture downloaded tweets as one dataset to NVivo 12 software and this enabled us to visualise the collected data.

Descriptive statistics were used to summarise the top 40 user locations, the most active accounts, and the top hashtag used alongside #DataSavesLives and identify the most prominent tweet (based on a number of retweets). We included both tweets and retweets in this analysis to get a broader picture of all Twitter users using the hashtag.

To understand attitudes towards the campaign aims among Twitter users, each original tweet was manually assigned a category as positive, neutral or negative attitude towards the campaign and big data research. The principles and techniques found in content analysis guided this process (Thelwall, 2013). We jointly created a short description of each category and then conducted a pilot coding of a sample of tweets during the team meeting. Based on these discussions, an experienced Twitter researcher (PT) systematically coded all remaining tweets.

Secondly, we undertook an inductive thematic analysis (Braun and Clarke, 2006; Braun and Clarke, 2021c). PT manually coded all tweets and the team met to identify, review and refine themes and choose the quotes representing them. Our research team is interdisciplinary, we work in and outside big data research, and one author (KF) is based outside the university, ensuring that we have both insider and outsider perspectives. Further analysis was carried out by mapping the thematic analysis results to the public engagement models, which offer insights into how Twitter users utilised the hashtag #DataSavesLives. Previous research has shown that the engagement techniques can be successfully mapped under these three engagement models (Metcalfe, 2019).

9.6.3. Ethics

The University of Liverpool Ethics Committee (9815) granted ethical approval. All captured data is publicly available online. Following established practice (Talbot *et al.*, 2020; Williams, Burnap and Sloan, 2017) when we use a direct quote, authors (excluding organisations) were informed and given an option to opt-out. No one asked to opt-out and one person requested a copy of published paper. We did not include pictures, links and emoticons.

9.7. Results

9.7.1. Descriptive statistics

Out of all tweets (n=3638) published in this period, the top 40 locations (excluding unknown) were from the UK, showing that the use of hashtag is still mostly based in the UK. Other countries included the USA, Australia, Germany, Spain, and Belgium (see Table 8) The discussion was dominated by professionals. Out of the ten most active accounts using the hashtag (which represents 48% of all tweets), all were non-individual accounts, such as organisations, networks or public bodies. All public body accounts were linked to the UK's NHS (See Table 9).

Country	Number of tweets
United Kingdom	2247
European Union (including Spain, Germany and Belgium)	76
United States	56
Australia	44

Table 8 Locations of Twitter users utilising #DataSavesLives

Twitter account	Number of tweets using #DataSavesLives (% of all tweets)	Type of organisation running the account
@hdr_uk	480 (13.2%)	Non-profit organisation
@usemydata	353 (9.7)	Non-profit organisation
@nhsx	261 (7.2%)	Public body
@nhsdigital	132 (3.6%)	Public body
@datasaveslives	125 (3.4%)	Non-profit organisation
@apha_analysts	97 (2.7%)	Network
@uk_healtex	85 (2.3%)	Network
@economics_unit	68 (1.9%)	Public body
@medconfidential	66 (1.8%)	Campaign group
@pioneer_hub	63 (1.7%)	Non-profit organisation

Table 9 The list of the ten most active Twitter accounts using #DataSavesLives

The most prominent tweet had 56 retweets and it discussed a new webinar on big data research and concerns around data privacy. Some organisations such as the Health Data Research UK regularly promoted the benefits of big data research using the hashtag (HDR UK, 2021).

Most of the hashtags used alongside the campaign were neutral or positive. The top ten included #healthdata (n=239), #covid19 (n=134), #nhs (n=102), #ai (n=101), #healtac2021 (n=91), #digitalhealth (n=89) #health (n=88), #testmining (n=84) #research (n=81), and #data (n=65) # The negative anti-campaign hashtag #DataGrab which was used by Twitter users accusing the UK government of trying to sell their medical data appeared nine times in the whole dataset and five times in original tweets. Thus rarely appearing alongside #DataSavesLives showing little cross-over between these two hashtags.

9.7.2. Attitudes

Discussion around #DataSavesLives was largely positive (n=716) or neutral (n=276) towards the campaign. There was some sarcasm in the negative attitudes (n=34) but no dark humour or personal attacks, which has been found in some other Twitter studies. This shows that the debate was generally conducted in a professional fashion, contrary to many politicised social media discussions (Teodorowski, 2020b; Karmegam and Mapillairaju, 2020; Hays and Daker-White, 2015).

Positive comments included reporting on successful, ongoing or future projects that had benefitted the public when using big data.

> *The University is partnering with experts from across the UK to launch a £2m data hub for mental health. The hub promises to speed up research into mental health and improve inclusiveness for disadvantaged groups #MentalHealth #DataSavesLives (*EdinburghUni*)*

This evidence of public benefit can be seen in examples of how big data helped the response to COVID-19 pandemic.

> *When the pandemic hit in 2020 we urgently looked at whether we could use routine data feeds to produce a more rapid cancer dataset that would help quantify the impact of COVID-19 on cancer services. This is one example of how that work is now being used #DataSavesLives @PHE_uk https://t.co/4Eu1QgxXGm* (EllissBrookes)

Twitter users often emphasised how important or relevant was their work around big data research. Thus, linking it to the campaign's underpinning rationale of showing that the reuse of medical data can change and even indeed save people's lives.

> *Our Hubs are working to improve health data so that researchers & innovators are better able to use it to enable discoveries that improve people's lives! #DataSavesLives Find out more: https://t.co/ZKQoaUWSos (HDR_UK)*

Often organisations would quote stakeholders, e.g. public members, to support these statements. There were calls for more public involvement and better data linkage.

Neutral tweets shared job opportunities, information about upcoming conferences, webinars or new publications, and asked people to participate in surveys or studies on big data research.

> *Hear from a super panel of speakers on Tues 25 May 10:00 -11:30 - A researcher's journey to accessing patient data. #datasaveslives #admindata (SCADR_data)*

Negative tweets did not always take issue with the campaign itself but raised concerns about the lack of public trust in the opt-out deadline for the new UK database scheme. Others picked up on wording used in hashtag and pointed out that the hashtag only appeals to professionals not the public and uses emotions to try to generate public support.

> *It's the wholly presumptuous nature of this scheme that is so abhorrent in my mind #DataSavesLives' the classic 'appeal to emotion' rolled out time and again as dogma in an attempt to upend logic #DataAsAsset is clearly much closer to reality* (griffglen)

9.7.3. Thematic analysis

We constructed five interlinked themes divided into five subthemes (Table 10) to illustrate how the debate around #DataSavesLives appears on Twitter. Figure 8 presents these key connectors and relationships between subthemes. We present the themes under the public engagement models of deficit, dialogue and participatory.

Themes	Ethical sharing	Proactively engaging the public	Co-producing knowleadge with public	Harnessing potential	Gaining an understanding of big data research
Subthemes	Trust & transparency			Excitement	
	Protecting individuals' rights			Space for improvement	
				Resonating motto	

Table 10 Themes and subthemes derived from the #DataSavesLives debate on Twitter through reflexive thematic analysis.

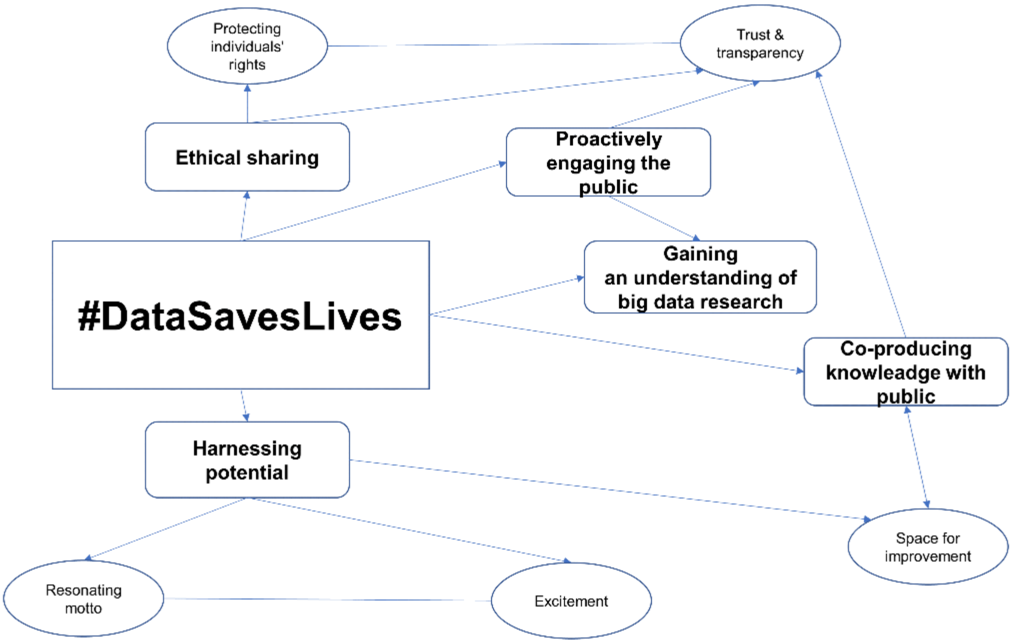


Figure 8 Key connectors and relationships between themes and subthemes.

9.7.3.1. Deficit

9.7.3.1.1. Harnessing the potential

Tweet authors on the whole thought that big data has the potential to offer benefits to patients, e.g. in the development of new medicines. Harnessing the potential of big data is visible in the following three subthemes: excitement, space for improvement, and resonating motto. The Covid-19 pandemic is present here but only as an additional argument for the claim that big data research is helpful for tackling new challenges.

9.7.3.1.1.1. Excitement

Tweet authors were often excited to announce new research projects and share study results (especially when showing how it has made some fundamental change or had the potential for real-life impact). Some of the tweets included authors who are passionate about the subject and others who were excited to participate in new studies. The researchers' success was recognised and noticed by the broader research community (e.g. receiving an award). Other tweets refer to upcoming events where authors were publicising their next presentation (this refers to both single events or conferences).

> *This is one of the most exciting pilots I've seen up close. How we can link patient data, what the analysis tells us and how we can provide evidence to make change for patient benefit.* #datasaveslives (SarahM_Research)

9.7.3.1.1.2. Resonating motto

Underpinning the campaign's motto is the argument that linking data and big data research saves and improves people's lives. This was a resonating motto, with many tweets about how the usage of medical data made an impact and provided new solutions. Tweets were either generic (relating to the benefits of big data research in general) or referred to specific research projects (both completed and ongoing).

> *'Data makes the unknown known' @margaretgrayson @useMYdata @NHSConfed #NHSReset #datasaveslives* (ConyersRebecca)

9.7.3.1.1.3. Space for improvement

Tweets also argued for some changes to ensure maximum benefit of big data. There were calls for more investment in big data research infrastructure showing that big data research is still developing.

> *Predictive data modelling could lead to better humanitarian outcomes, but we are missing half the data needed. Time to act! #DataSavesLives*. (Enovacom_en)

Twitter users also recognised that some of these changes had to happen soon to offer more benefits from research.

> *Ahead of a crucial @G7, @NMRPerrin argues for the urgent need for better coordination across the global data sharing landscape https://t.co/aw8Apgw5Ku #datasaveslives @GS_Humphreys @royalsociety @GloPID_R* (ICODA_research*)*

9.7.3.1.2. Gaining an understanding of big data research

This theme is about reaching others (including the public but primarily other professionals, policymakers and researchers) and offering an opportunity to learn more about individual projects.

The hashtag offered an opportunity to call people to action, to apply for job openings (mostly research-related) and welcome new team members. Some tweets asked other researchers to support big data research or answer ongoing consultations or surveys.

> *Only a few days left to apply for this! Working with a great team enabling the #HealthData infrastructure to support #COVID19 #research. Secondments welcome, remote working too so location flexible. #HealthData #DataSavesLives (LaraEdw001)*

This illustrates how the hashtag was utilised among similarly minded people to publicise new opportunities and events.

Tweets also allow readers to learn more about big data projects, attend events, follow online chats and read recent blogs or papers. This is mostly passive and focused on dissemination rather than engaging.

> *Check out this thread from @HDR_UK with examples of how #DataSavesLives being added throughout June* (NIHRresearch)

9.7.3.2. Dialogue model

9.7.3.2.1. Ethical sharing

The need for ethical, safe and lawful sharing of data in big data research and the importance of doing it right was a prominent theme in the data. Two subthemes deal with key aspects of achieving these aims: protecting individuals' rights, and trust and transparency.

9.7.3.2.1.1. Protecting individuals' rights

There is agreement that big data research offers new opportunities for innovation. However, the impact on individual rights remained the main concern. This was particularly around how the data is used, who has access, how secure it is, and if patients could be identified. Many organisations attempt to reassure people by telling them that any data usage is secure and transparent.

There were concerns that health data could be sold to private companies to make a profit. Some tweets linked that concern with people's decisions to opt-out in the UK from using their medical data for research. Some admitted that the public has not been properly or sufficiently engaged around and about these issues.

> *@Axelheitmueller, you're completely correct, the benefits of data sharing are immense for the health of our nation. For some reason there's a narrative that we intend to make a profit from data. This is simply not the case. We do not, and we will not sell data!*
>
> *#datasaveslives (simonrbolton)*

9.7.3.2.1.2. Trust & transparency

Associated with individual rights is trust and transparency, which underpins public support for big data research. Tweet authors argued that public trust is essential for big data research to succeed and that the processes of data sharing have to be transparent and follow well-established principles. Otherwise, it risks undermining public support as the public will lose confidence. There have been comments within the UK context that recent political events have undermined that trust, which is also shown by the hashtag #DataGrab. Trust and transparency are perceived as the building blocks of successful research projects and is often the rationale that underpins public engagement.

> *Sharing my data can aid research needed to improve healthcare for myself and others with chronic illness. However, there does need to be clearer reassurance that data won't be misused so that individuals can make an informed choice. #GPDPR #nhsdataoptout #DataGrab #datasaveslives* (LucindaH19)

9.7.3.2.2. Engaging the Public

There was a push in the tweets to have better engagement with the public and encourage conversations about big data research. Some approaches to this included avoiding jargon and ensuring that events are free to attend. There was also some media engagement as Twitter users shared links where researchers taking part in media interviews. Also, media outlets were tagged as Twitter users tried to catch their attention. These engagement activities are intended to help the public to understand the value of big data research better. However, if they limited themselves to only explaining big data research to the public, they could be seen as following a deficit model of engagement, with its associated limitations.

> *Health data research can be confusing sometimes and full of buzzwords and jargon. This article clearly explains how health data is used and why it's so important. If you donate your data to health research you could help improve future healthcare. #DataSavesLives #DataScience* (genscot)

9.7.3.3. Participatory model

9.7.3.3.1. Co-producing knowledge with public

Public contributors could be successfully involved in big data research. These are public members who actively contribute to the research projects ensuring that research is done *with* and not *to* or *about* them. Views on how much the public should be involved differed. Some tweets explore the active role of the public in studies as public contributors, whereas others focus only on reaching people and showing them the benefits of big data research (as shown in the previous theme, proactively engaging the public).

Tweets refer to involving public members in big data projects. In this theme, there is a call for more public involvement. Tweet authors show examples of how involving public as active contributors had a positive impact on their research.

There were calls for: more public control; thanking patients for sharing their medical data for research (not opting out); and recruitment calls for new public contributors in big data projects.

> *"None of this would be possible without our Data Trust Committee – the diverse and inclusive group of patients and members of the public, who review every data access request and make decisions based on the Five Safes and, ultimately, the public's best interest." #datasaveslives* (useMYdata)
>
> *It is also important to involve patients in developing registries or data collections. Also citizens, because they produce the data and therefore, as owner of the data, they should have a seat on the "Datatable" too. #patientsinvolved #datasaveslives #MTF2021* (Birgitpower)

9.8. Discussion

This study explored how #DataSavesLives was used on Twitter. The findings clearly show that the debate was mostly positive towards the campaign. This is not surprising as most participants were organisations, academics, and institutions that work in big data research. Our findings confirm previous research on the #DataSavesLives hashtag; that it is being used to identify similar-minded projects around big data and to spread positive messages towards big data research particularly when there are public controversies (Hassan, Nenadic and Tully, 2021).

We mapped the results of our thematic analysis into models of public engagement. This has shown the largest number of themes were within the deficit and dialogue models and only one theme was included in the participatory model. Each model has its uses, and a hierarchy is not necessarily the most useful way to understand them (Trench, 2008) and the public engagement campaign can be placed within all of these models (Bucchi, 2008). However, if the campaign wants to improve trust with public members, more active exchange with the public is needed. This can be achieved by moving more campaign-related activities into activities that would conform with the dialogue or participatory models. One way of doing this is to engage more Twitter users to participate in active discussion online. Previous research has shown that Twitter can accommodate a vibrant debate around challenging topics (Richardson *et al.*, 2016). How Twitter users utilised the hashtag #DataSavesLives is not a new phenomenon in Twitter discussions about science. For example, a study that explored science festivals found that organisations mostly focused on distributing information and only smaller part of the twitter activity was actually interactive (Su *et al.*, 2017).

The hashtag usage remains limited to similarly-minded Twitter users – a Twitter bubble. The results indicate that communication around big data research is driven by the professional community and research remains one-way because the public rarely uses the hashtag. This confirms previous research showing that government science organisations do not fully utilise the potential of social media to engage with the public (Lee and Vandyke, 2015). Within this dataset, there was only a limited appearance of negative hashtags such as #DataGrab (n=5), which was used during the UK debate on the new database scheme. This elicits questions about how successful the campaign is in achieving its goals of engaging with public. The campaign messages do not target any seldom heard communities but rather focuses on researchers and professionals. Twitter bubbles are not a new phenomenon, and Sunstein (2018) describes it as an “echo chamber” which amplifies the already existing beliefs of Twitter. However, despite the public not using the hashtag themselves, it does not exclude the possibility that they are exposed to these messages, as research (Côté, Darling and Heard, 2018) has shown that researchers with over 1000 followers on Twitter have diverse followers (including media representatives and public members). The #DataSavesLives campaign shares many aspects of one-way communication and remains in the deficit engagement model. However, many engagement campaigns have limited interaction with the public at the beginning but can improve over time (Bucchi, 2008). Thus, based on previous research the campaign has potential to develop.

The campaign was relaunched in Europe in 2019, but there were only four Tweets in languages other than English. Our findings indicate that the campaign remains UK centric as the most active Twitter accounts are based in the UK. The high activity of the government-run UK organisations poses the question if the hashtag and campaign could continue on Twitter without their involvement. The use of #DataSavesLives remains limited on Twitter. However, this can be explained by the type of messages published online. Most were positive or neutral towards the campaign, whereas the negative emotions on social media spread faster than positive (Zhu, Kim and Park, 2020). This should not encourage Tweet authors to start appealing to negative emotions but rather recognise the limitations of the positive engagement campaign.

Ethical challenges, issues of trust and transparency around big data research remain a concern for the public (Stockdale, Cassell and Ford, 2019). In 2014 NHS England launched a promotional campaign showing how medical records would become a part of a larger database. The project called Care.data was controversial and a previous study explored the #caredata controversy on Twitter (Hays and Daker-White, 2015). At that time there was a distinct lack of public engagement or involvement in big data projects. There now seems to be a clear recognition that the public should be proactively engaged and involved in discussions about big data projects. There is an improvement in how professionals and organisation perceive public involvment. According to Tweet authors the public can be involved at various points (Arnstein, 1969). Some suggest only explaining the benefits of big data research, and others call for and present examples of having public contributors involved in research (e.g. governance). Limited public understanding of the use of big data remains one of the largest challenges (Skovgaard, Wadmann and Hoeyer, 2019), and more engagement could, arguably, improve this situation.

Based on our research findings, PT participated in a Tweet Chat hosted by the European Patients' Forum as a part of their regular conversations around big data research on Twitter. We hoped that this would allow more online engagement within the dialogue model. The discussion considered the online movement and how social media is spreading the campaign's message (*Reflecting on the online movement – how social media is spreading the message of Data Saves Lives*, 2021). We found it beneficial to present our research, discuss the emerging findings and engage with Twitter users who had used the hashtag #DataSavesLives. This was an opportunity to talk to the people involved in running the campaign about what they thought the future of the campaign might be. The public member contributing to the discussion pointed out the need for more actively involving public around big data research. This further confirmed our findings and the need for researchers to shift engagement to dialogue and participatory models.

9.9. Limitations

Organisations in the UK were the main authors of downloaded Tweets. This limits our understanding of how much the results of our study reflect public attitudes towards the campaign and questions if the public is actually aware of it. Twitter offers limited demographics about its users. Some data, such as location, was unknown (e.g. online location appeared as the third most popular location, used by 7.6% of Twitter users) or included two or more countries. Also, because some demographic data was unavailable, we cannot say if the usage differs among different age groups or other attributes.

The activity of an automated Twitter account, a bot, can influence Twitter traffic. A bot aims to create tweets and retweets to expand the coverage of their messages. We manually coded the dataset and did not notice this kind of activity, but this does not guarantee that it was not there.

Data collection took place when there were new database scheme controversies in the UK, which could have influenced some traffic and messages. Future research should check if the Twitter discussion has shifted depending on the context. Our study explored only usage of #DataSavesLives in English, but it is also available in German as #DatenrettenLeben. Our study focuses on Twitter, the main microblogging platform where users often discuss contentious and/or political topics. However, the hashtag is also available on other social media (Facebook and Instagram), future research could explore if engagement there differs from Twitter. Other research could also focus on negative hahstags towards sharing routinely collected health data such as already mentioned #DataGrab.

9.10. Conclusions

This study shows how Twitter users utilise #DataSavesLives when the hashtag was adopted by the UK government and during the UK domestic controversies around data linkage and sharing. There are growing expectations from funders that researchers will engage with the public. Social media campaigns such as #DataSavesLives may offer the opportunity to further this goal. This study expands our understanding of the #DataSavesLives campaign. The results demonstrate the potential of social media and recognises the need for engaging with a wider range of opinions and different Twitter constituencies. Thus, researchers need to identify new ways of actively engaging a wider range of the general public. There is a need to move engagement activities from a deficit model to dialogue and participatory models that include active two-way between researchers and public members, and genuinely involving the public is meaningful. Future research could explore if and how Facebook and Instagram users use the hashtag.

10. Researchers’ study

This chapter shows the findings from the second primary study as published verbatim in the Health Expectations titled *‘To me, it's ones and zeros, but in reality that one is death’: A qualitative study exploring researchers' experience of involving and engaging seldom-heard communities in big data research* (hereafter “researchers’ study”) (Teodorowski *et al.*, 2023a). Prior to presenting the paper, I introduce it by discussing how this study links to the literature review. Then, I briefly explain the study's aims and how the findings contribute to the current body of literature.

10.1. Linking the literature review with the paper

The relevant issues from the literature review for the researchers’ study were:

- There is limited understanding of how to involve and engage seldom-heard communities in big data research.
- No empirical studies explored the experience of people who involved and engaged members of the public around big data research.
- There is no one standardised approach to involve and engage members of the public in big data research.
- Involving and engaging members of the public in big data research could include additional challenges.

The approach in this study was to gather the experiences of researchers who involved and engaged seldom-heard communities around big data research and thus address gaps and issues identified in the literature review. This was achieved by conducting semi-structured interviews with researchers based in England, Scotland, Belgium and Canada. As different communities can be considered seldom heard, participants could have decided for themselves if the community they worked with was seldom-heard. This broad approach was appropriate as no prior studies looked at this topic.

The study explored the involvement and engagement process in big data research from the perspective of researchers. The findings indicate this could be conducted meaningfully for researchers and members of the public. Lessons from this study could also benefit researchers in other health research projects (outside big data) on how to work with seldom-heard communities, as the study discusses barriers and enablers for working with these groups. The findings emphasise the importance of power-sharing between researchers and public contributors. Findings confirm previous literature that there is no one way to involve and engage members of the public around big data research but also provide further examples of how this could be done. Based on the research, the paper suggests key recommendations for involving and engaging seldom-heard communities in big data research that could overcome some of the challenges. I explore these further in the discussion chapter.[57] Last, the findings indicate that researchers should consider which community would be seldom-heard in their work.

10.2. Abstract

Background

Big data research requires public support. It has been argued that this can be achieved by public involvement and engagement to ensure public views are at the centre of research projects. Researchers should aim to include diverse communities, including seldom-heard voices, to ensure a range of voices are heard and that research is meaningful to them.

Objective

We explored how researchers involve and engage seldom-heard communities around big data research.

Methods

This is a qualitative study. Researchers who have experience of involving or engaging seldom-heard communities in big data research were recruited. They were based in in England (n=5), Scotland (n=4), Belgium (n=2) and Canada (n=1). Twelve semi-structured interviews were conducted on Zoom. All interviews were audio recorded, transcribed and we used reflexive thematic analysis to analyse participants' experiences.

Results

The analysis highlighted the complexity of involving and engaging seldom-heard communities around big data research. Four themes were developed to represent participants' experiences. These were 1) abstraction and complexity of big data, 2) one size does not fit all, 3) working in partnership, and 4) empowering the public contribution.

Conclusion

The study offers researchers a better understanding of how to involve and engage seldom-heard communities in a meaningful way around big data research. There is no one right approach, with involvement and engagement activities required to be project specific and depend on the public contributors, researchers' needs, resources, and time available.

---

[57] See at page 227.

Patient and Public Involvement

Two public contributors are authors of the paper and they were involved in the study design, analysis, and writing.

Keywords

Public involvement; seldom-heard; public engagement; big data; PPI; qualitative.

10.3. Introduction

Patient and public involvement and engagement (PPIE) has become embedded in health research and within the NHS (NHS, 2020), and is used in healthcare services (Mockford *et al.*, 2012) to put the public perspective at the centre of the discussion (Sofolahan-Oladeinde *et al.*, 2017) and improve professionalism among medical practitioners (Lalani *et al.*, 2019). It helps to align priorities shared by researchers and the public (Aitken *et al.*, 2019) and it helps researchers understand the lived experience of patients and the public (Glasby and Beresford, 2006). There is also an ethical argument that those who pay (taxpayers) should have a say on how their services and research are shaped (Tritter, 2009). We follow the National Institute for Health and Care Research (NIHR) definition of public involvement and engagement (NIHR, 2021a). Public involvement in research means that work is "*being carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them*." We use the term "public contributor" to describe this role. Conversely, public engagement stands for activities "*where information and knowledge about research is provided and disseminated*."

10.3.1. Big data

There are multiple definitions of big data in the literature (Mehta and Pandit, 2018). In this paper, we define big data research as reusing routinely collected medical data for research purposes. This can happen by linking large medical datasets from various sources. When initially collecting medical data, the public (or the researcher) might not be aware that their data may be later reused for research. Many big data research studies use opt-out consent, where patients need to inform someone, usually their medical provider, that they do not want their medical data to be reused for research.

Public support is needed for these projects to be able to take place (Taylor, 2014), and a systematic review has shown that the public generally supports the reuse of their medical data (Aitken *et al.*, 2016). However, they can be concerned that their data might be misused, e.g. sold to private companies (Hays and Daker-White, 2015). PPIE can assist in alleviating these concerns (Muller *et al.*, 2021). Hill and colleagues found that talking about and explaining the research process around big data improved their study participants' support in reusing their medical data (Hill *et al.*, 2013). Public contributors can also contribute to the decision process on who can access medical data for research purposes, thus ensuring that a social licence exists (Ford *et al.*, 2019). Social licence is more than meeting legal requirements and requires public trust that researchers will conduct their work ethically (Muller *et al.*, 2021). Poor governance can lead to a deterioration of the social licence (Dixon-Woods and Ashcroft, 2008).

10.3.2. Seldom-heard communities

In addition to the "usual" public, it is important to capture the voices of groups in our communities who are less frequently heard. Successful PPIE requires the inclusion of seldom-heard communities, (Harrison *et al.*, 2019b; INVOLVE, 2012a; Aitken *et al.*, 2019) and researchers should aim to include them, (Dawson *et al.*, 2018) but how to do it in a meaningful way remains challenging (Flynn, Walton and Scott, 2019; Ocloo and Matthews, 2016; IPFCC, 2018). Such communities are often easy to ignore, but not including them can make research findings ungeneralisable to all parts of society and miss the nuances of experiences specific to those groups (Bonevski *et al.*, 2014) and will not provide solutions for all communities (Madden and Speed, 2017). PPIE should be inclusive and accessible to everyone (Aitken *et al.*, 2019). Not including seldom-heard voices can reflect the power structures at play and perpetuate health inequalities. This is important as these communities might experience poorer social and health outcomes. For example, the Covid-19 pandemic disproportionately affected people from ethnic minorities (Kumar *et al.*, 2021).

The terminology and definitions in this area are contested. Some of the terms used include: hard-to-reach (Bonevski *et al.*, 2014; Nguyen, Cheah and Chambers, 2019), seldom-heard (Rayment *et al.*, 2017; Prinjha *et al.*, 2020), seldom-listened (Tierney *et al.*, 2021), peripheral voices (Hanafin and Lynch, 2002), marginalised (Silva, Smith and Upshur, 2013; Snow, Tweedie and Pederson, 2018) and under-served (NIHR, 2020). The key characteristic of these definitions is that these communities are less included in research than other groups in mainstream society. Within the UK legal context, the Equality Act 2010 uses the term 'protected characteristics'. These are age, disability, gender reassignment, marriage or civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation. The Act provides anti-discrimination laws and embeds requirements for diversity and inclusion for public bodies but is not always directly applied to research. But it can be influential in how researchers approach diversity in their work (INVOLVE, 2012a). We will use "seldom-heard" as this shifts the responsibility for inclusion to researchers rather than blaming the public, as implied by the "hard-to-reach" wording. Their inclusion (or a lack of it) is not a fault of these communities (NIHR, 2020). When presenting the results, we kept the original terms used by participants when quoting them. However, we recognise that use of any terms might not necessarily represent how these communities would like to be described.

10.4. Research aim

Despite understanding the importance of PPIE, there is limited knowledge of how this can be effectively facilitated in big data research (Teodorowski *et al.*, 2021a). A previously published system logic model identifying key elements of PPIE in big data research recognised the inclusion of seldom-heard communities as a key component, (Teodorowski *et al.*, 2021a) and therefore there is a need to understand how to ensure all voices are included.

This paper explores researchers' experiences of involving and engaging seldom-heard communities in big data research.

10.5. Methods

10.5.1. Theoretical position

This study adopts social constructionism as its theoretical lens when understanding and analysing data (Gergen, 2015). We believe that multiple realities and perspectives exist among researchers. These are subjective and socially constructed and thus depend on participants' cultural, political and historical backgrounds. Researchers (and thus their work) are shaped by their relationships with public contributors. From the social constructionism perspective, the dynamics of social interactions are essential to understand how new knowledge is achieved (Burr, 2015). Thus, in our analysis, we focused on the processes around PPIE rather than its structures.

Social constructionism can be used to justify a more collaborative form of inquiry (Gergen, 2015). This can be achieved by conducting research together with the public contributors. Collaborative work can be seen among our participants who involve the public in their work, but also in our project as we involved two public contributors as co-researchers.

10.5.2. Participants and data collection

Alongside big data researchers, we included facilitators of PPIE activities in big data projects. Facilitators (some of whom might be qualified researchers) are in charge of the overall organisation of the PPIE progress, they co-ordinate, organise and facilitate activities and act as intermediaries between researchers and public contributors (Todd, Coupland and Randall, 2020). They often are recruited at research institutions to support specific big data research projects. Throughout the paper, for clarity, we will refer to both groups as researchers. All participants had to have an experience of involving or engaging seldom-heard communities or aiming to reach them.

We recruited through Twitter, bulletins and established networks within big data research such as Health Data Research UK. Interested participants contacted the author for further details and to register their interests. Interviews were conducted on Zoom between March and June 2022. Interviews were later transcribed and anonymised with all participants assigned pseudonyms.

A semi-structured interview guide was developed to elicit participant experiences of PPIE with seldom-heard communities. We also included an opportunity for them to speak about communities they planned to reach or tried to engage but were unsuccessful. After the first interview, co-authors met to reflect on the topic guide. One follow-up question on what participants perceived as a seldom-heard community was added to the topic guide.

Only limited demographics were collected from participants to protect their anonymity. Twelve participants took part in the study. We reached data saturation when no new themes appeared in our analysis (Guest, Bunce and Johnson, 2006). Participants were based in England (n=5), Scotland (n=4), Belgium (n=2) and Canada (n=1). The majority were women (n=11) and one man. Their experience of research and PPIE was between two and a half to twenty years, with an average of nine years. We also asked them to describe themselves as researchers (n=6) or facilitators (n=9), although they could have chosen both options. Six participants were from an ethnic minority background.

10.5.3. Data analysis

We conducted a reflective inductive thematic analysis (Braun and Clarke, 2006; Braun and Clarke, 2021c; Braun and Clarke, 2019). This method allowed us to identify patterns across all interviews systematically. Thus, we unpacked the realities experienced by researchers. We used both semantic and latent coding. Semantic coding shows more explicit patterns within the data and stays as close as possible to what participants said. This allowed us to share specific practical examples of PPIE strategies. Latent coding provided more implicit and interpretive reflection on the data. Initially, one interview was coded jointly by three authors (PT, SA, NT) in Word. Then, the author (PT), an experienced qualitative researcher, inductively coded the remaining interviews, supported by NVivo 12. We met as a team on multiple occasions to discuss the data analysis, and develop and refine further themes.

Public contributors can be meaningfully involved in qualitative analysis (Garfield *et al.*, 2016) and trained to conduct reflexive thematic analysis (Hemming *et al.*, 2021). Two authors (SA and NT) who are public contributors received training in reflexive thematic analysis (focusing on being reflective, coding process and refining themes). They were involved in the study design, initial coding and developing and refining the themes. They are both experienced public contributors and SA also acts as the Data Ambassador for Care and Health Informatics theme within Applied Research Collaboration North West Coast. This role involved raising awareness and knowledge about big data research. Two authors (SA and NT) also took part in an exercise reflecting on how their backgrounds influenced what they perceived in the data and what they brought to the analysis.

The research-active authors also reflected on their academic backgrounds (PT, KF, SER and LF). Those who were qualitative researchers with experience of involving and engaging the public, and those who conducted research in big data. These different perspectives allowed us to bring distinct views to the data analysis and furthered our understanding of the experiences of our participants.

10.5.4. Ethics statement

We received ethical approval to conduct this study from the ethics committee at University of Liverpool under the number 10063.

10.6. Findings

We present four themes that explore how researchers involved and engaged seldom-heard communities in big data research. These are: 1) abstraction and complexity of big data, 2) one size does not fit all, 3) working in partnership, and 4) empowering the public contribution. All themes appeared throughout all interviews which gives an indication that these experiences were commonly shared among participants (even if participants were based in different countries). We provided additional quotes in Appendix 9 that offer further examples of how the participants involved and engaged seldom- heard communities around big data research.

10.6.1. Abstraction and complexity of big data

Big data can be an abstract and difficult topic to explain to the public. Participants said that conversations about big data include technical, specialist's vocabulary, jargon, references to legislation and regulations. Researchers found it challenging to discuss the complexity of this kind of research with public contributors in lay terms:

> "*Big data is a really complex environment to navigate both in terms of the research, but also in terms of like the regulatory aspects and legislative aspects.*" (Sophia)

Sometimes, the difficulty in explaining big data research impacted on participants' experience of involving the public. Public contributors can have a role in advising (or deciding) if researchers may access routinely collected health data for research purposes. Here, the public contribute to the governance groups of these initiatives. Researchers who worked with these groups found it hard to explain to the public the purpose of big data research. They struggled to contextualise the concept of big data to the public if it did not directly refer to the public contributors' health condition or a topic that might interest them. The following extracts illustrate that challenge as the participant refers to bringing public contributors to support big data infrastructure:

> "*Project (…) was just looking at the infrastructures of big data. It was really challenging to actually put that into a context that was relevant to members of the public; they kind of said 'well we don't even know if you want us to be involved, we don't really see how we can be because this is all to do with linking up datasets with each other and it's all very technical, and it's not really anything to do with our living experience as patients or as members of the public'. So that was that was quite a hard project actually to think about*." (Sienna)

It is not only public contributors who can be confused by big data jargon. Some participants who were not data researchers said that their familiarity with the topic was more akin to the public contributors rather than data researchers they worked with on the project. They might feel uncomfortable asking questions or requesting clarification. The public contributors often were more confident in asking these kinds of questions. This was seen as a very positive element of PPIE by a participant:

> "*I'm sometimes really pleased when [public contributors] ask questions. Because I'm like oh, good, I don't know if I could have asked that, but so I'm really pleased that you did. I probably should have known that, but I don't, so I'm glad you asked it*." (Robyn)

Participants felt that promoting the benefits of big data research, being transparent in how data is used, and building trust with the public would ensure that some negative media stories around big data research could be counteracted. They believed that overall, the general public would be supportive of data sharing to improve healthcare. They recognised the need for effective communication between researchers and the public. In individual projects, they suggested training and supporting the public contributors around big data research but described it as a slow and time-consuming process.

> "*One of the things that we really do is kind of work with our staff to make sure that they are able to explain it in kind of like plain English. If we were to have a session about something like trusted research environments, which can be kind of like a technical. Then we would work with staff to actually plan the presentations (…) to make sure that the language is right, we also hold drop-in sessions once a month so that members of the public that we work with can come in and say 'I have a question'. (…) And so we*

> *bring in some of our more technical staff because I have no technical knowledge myself*." (Harriet)

Participants spoke about how communication must continue outside the research projects and involve the broader community. The public contributors involved in big data research are essential to helping further engagement with their communities. As they become more familiar with big data research, their knowledge can be utilised to engage with the general public and raise awareness of big data research. They can help explain what big data research is about, its benefits, and how it works. Here, a participant speaks about explaining in lay terms a technical term related to data:

> "*When it comes to data and infrastructure and things, it can be very complex. There's lots of big words like pseudonymisation [laughing] and things like this, so we worked with the public members to create this animation, which gives a snapshot of what the project's like and it's an accessible snapshot*." (Robyn)

This theme shows that talking about big data can be complex and challenging. However, there was an agreement that PPIE around big data research takes the researcher away from numbers and allows them to bring a human face to the data. This excerpt explains this:

> "*I love doing this type of analysis of, you know, hitting the buttons and seeing the graphs come up and seeing results. It's really exciting, but you miss that contact with people. And having that PPI group, there was a really good way for me to touch base and think about what the numbers meant. And think about the stories behind some of the data. And connect it to people's lived experience and I think that's really important. (…) To me, it's ones and zeros, but in reality that one is death. So it's really important to have that in front of your mind, and I think that brings it home when you've got a group of people in front of you who are really interested in what you're doing and to whom it could potentially make a difference*." (Zoe)

10.6.2. One size does not fit all

This theme elicits the need for researchers to be flexible and often innovative when involving public contributors in big data research. Participants did not have one prescription on how to successfully work with the public contributors.

How PPIE looked in the participants' work differed based on the project needs, public interest, or experiences. Public contributors can be involved in different roles within projects around big data research. These included contributing to the review of the data access process, and as co-investigators or being members of advisory groups for specific projects. The following quote shows how public contributors can assist with decisions over whether and how researchers can access routinely collected medical data for research purposes.

> "*That's a group of around eight members of the public who we meet with on a quarterly basis to get their views on our kind of engagement plans (…) and also to get them to become more part of our project approval process is something they've been really keen to do, so we're looking at our kind of review process. Researchers who want access to routinely collected health and social care data puts their applications in and it goes through a rigorous, multistage approval process and one of those that we're looking to do is to have the public voice within that so their vote, their part of it would be an assessment of the public value of the projects that come in.*" (Alex)

Participants said that public contributors can have a much more active role and co-share responsibilities with researchers:

> "*We have two co-leads. One of them is myself and but the other one is a member of the public, so that from the very beginning, I am working very closely with [the public contributor] so that we can kind of shape this programme together, making sure that the public views are fed in right from the very beginning and as part of that we've also got a leadership (…) and so in this leadership team, it's half public contributor, patient-public contributors and the other half would be kind of like professionals such as myself*." (Harriet)

How to work with each community might depend on their needs. Many participants spoke about the need to understand the specific community they were planning to work with. Here, a participant suggests a pre-engagement engagement to understand what PPIE should look like:

> "*It's just really interesting about doing that pre-work to set up the scope and the scale of the engagement work and then to set up the environment that would be the safe as possible, so it's almost like a pre-engagement engagement where you're really setting up the safe environment to allow for good public engagement to happen for diverse members.*" (Victoria)

Who represents seldom-heard communities differed among participants. Participants often spoke about aiming to be reflective of the community. However, they recognised that it was not always possible (or feasible) to reach everyone who might potentially contribute. They admitted that because of their recruitment methods, limited resources or time, the public contributors who were generally involved often represented a limited range of demographics.

Each community is different and might require different PPIE strategies. They argued that the recruitment should be specifically tailored to the group they wanted to reach. The communities that were most often involved in PPIE were generally white and elderly. The seldom-heard communities they wanted to involve included ethnic minorities, people experiencing homelessness, traveller communities, or different age groups (especially younger people). However, they also wanted to reach people with particular health conditions or improve male representation. This quote illustrates how participants perceived their role in encouraging diversity:

> "*We do try to reach out to seldom-heard groups. We are currently undertaking an audit of our group to see how, where we're lacking, 'cause I suppose within the patient and public involvement there tends to be a certain type of person who volunteers and has got the time. So tend to be retired, tend to be white more often than not, and so we are keen to widen our demographic (…) we're not just interested in ethnicity (…) it tends to be quite a lot of women as well that volunteer, so you know, increasing, men, also increasing our younger population.”* (August)

10.6.3. Working in partnership

PPIE is not conducted in a silo. The participants worked with others (organisations, charities, public services, and public contributors) with the aim of being inclusive and to reach more diverse communities, especially around big data research. This theme explores these different actors' roles in successful PPIE.

These partnerships have the potential to fill the gaps in researchers’ understanding of local communities. Some participants recognised that researchers themselves could be a hard-to-reach group. Meetings can be held during working hours or be otherwise inaccessible to public contributors. Others recognised that the diversity of their teams is important and might reflect how well they involve and engage communities.

> "*I think while we don't have as much diversity as we could in our staff, it's harder for us to communicate or share those messages or understand the groups that we're trying to reach."* (Arabella)

Charities and organisations already provide existing links with the community and offer that bridge for researchers to reach the seldom-heard groups. They can assist with recruitment and engagement strategies. However, there is a risk that a researcher will not necessarily improve the diversity of their group but rather take over the demographic composition of the group they engaged with, as this participant explains:

> "*So it was mainly about because I was kind of piggybacking on a charity, on several charities groups. It was down to who they had picked up and they were already actually meeting via Zoom this charity, so I kind of inherited their diversity or degree of diversity*." (Zoe)

However, as much as these partnerships can be helpful, establishing them is not easy. It can be time-consuming to build that trust with the charity, and participants recognised that this needs to be an ongoing relationship that should benefit both parties.

Some participants also said that that relationship could be confusing to the potential public contributors if there is more than one research team working on that project (and thus trying to involve them). The following extract shows how one of the participants struggled to get some patient groups involved because they already had been working with other researchers:

> "*I contacted several [patient groups] in [the city] to see if they would be interested in doing some PPI workshops with them or telling them a bit more about the research we're doing. (…) They didn't necessarily know that they it was the right thing for them at the time, but also they'd had so many researchers getting in contact with them that it's they said it's just really difficult for us to choose who we work with and if they've already got a relationship with somebody else. Then they may choose to work with them obviously instead*." (Sienna)

Researchers can act as facilitators of PPIE or bring in trained experts (who might not necessarily be familiar with big data research). The facilitators' role is to act as this connecting bridge during work, an intermediary between researchers and the public contributors.

> "*What we are trying to do is bring these people on board and explain to us what it is, and we try to turn it into more lay language and sometimes with [public contributors], engage them to have a conversation so that they can actually challenge the experts rather than us doing it. So we are more of an inbetweener in that sense*." (Kimberly)

PPIE is also about involving individual public contributors. Participants often spoke about how interested and passionate public contributors can become about their involvement. These partnerships require working together and respecting each other. Some participants spoke highly of public contributors they worked with:

> "*And one thing that I think that is often forgotten is about [public] members is that they are just, they're not just patients or they're not just a member of the public. These are very talented, very skilled people. You know they've got their own life skills. You know they've got their own careers. They've got all of the skills and knowledge from that, and*

> *I think it's great that they want to volunteer with us and help share some of that*." (Robyn)

Only when truly working in partnership with public contributors can it lead to their empowerment. This is the focus of the next theme.

10.6.4. Empowering the public contribution

Participants felt that for involvement to be successful, there must be a power balance between researchers and the public contributors. Empowerment gives public contributors the ability to contribute to the involvement process fully. This can be achieved through ongoing support and ensuring that they become more familiar with big data research or projects they are involved in. As a participant illustrated below, this is a continuing process.

> "*Giving a sort of chance for people to ask questions, which was the nice thing about that project is that it wasn't a one-off, people could go away, look up something for themselves and then they could come back and be like what's this and they'd post a link and then we'd come back and answer those questions. So it was quite a nice kind of two-way in that sense*." (Drew)

Most participants felt that public contributors need to be supported at each stage of the involvement process but also recognised that this can be time-consuming and require addional work. Some suggested an open-door policy where public contributors could reach researchers anytime and thus also feel like a part of the team. WhatsApp groups for public contributors can be a safe place to discuss the project further. Public contributors should receive training or induction both around the project and PPIE (especially if they are involved in a research project for the first time). One of the techniques which supported the public in understanding the jargon around big data research was a 'live dictionary' which could be updated as people asked questions throughout the lifetime of the research project.

> "*But one of the things that we've created is an ongoing glossary. And if there's any words or phrases that the [public] members don't understand, it's a case of pop it into that glossary, and someone will answer it for them*." (Robyn)

However, participants recognised that not all training can be equally helpful and that some institutional resources were more bureaucratic and could potentially discourage people from being involved. This is illustrated by the following quote talking about the focus on training offered by the academic institution to new public contributors involved in the research:

> "*[The training] is quite formal and it's about like the whole university obviously it's not about big data, it's not really keyed towards seldom like heard groups or different types*

> *of groups, so I think there's other types of training that could still be useful for people, even if it's just, you know, stories of being involved that are from people who are more like them. So I think it could be a little bit of a little bit tailored, and some of it's very dry if I'm honest*." (Zoe)

After receiving all this training and support, some participants felt there is a danger that the public contributors start offering more of an expert view rather than a lay person perspective. There is a fine balance between understanding the project enough to be able to provide a nuanced contribution and where public contributors become what one can describe as 'usual suspects' of people who keep being involved and thus become more like professionals. One participant spoke of a successful approach to dealing with this challenge:

> "*it is a really fine line between building their knowledge to get involved and becoming an expert in that and kind of losing that public perspective (…) to kind of help with that; we do also have members of the public in a role for only specific amount of time. So, for example, now [advisory board]. They're only there three years, and then we kind of refresh the board, so with that, we're constantly bringing in that kind of like newer public perspective as well*." (Harriet)

Empowerment must be felt in practice and involvement needs to be genuine. Public contributors must feel that they make a difference. In the 'one size does not fit all' theme, a researcher spoke about the public contributors' panel assessing if researchers can access medical data for research purposes. The participant described how the public contributors perceived this and how it could be expanded for more empowerment:

> "*'Do you agree with our decisions over whether these were approved or not?' And in the main, they aligned with what the decisions had been, but on a couple of occasions, they were like 'we don't see the public value in doing this. It's not well explained', so is either it wasn't when explained or the public value wasn't there, and so that going more of a point of challenge for them and made it quite clear that they wanted to be part of the genuine process of review*." (Alex)

Participants pointed out that only when there is a real sense of empowerment can public contributors' involvement impact positively on the research projects. Multiple ways exist of how public contributors can shape projects. Participants named the following contributions: ensuring the research questions address the public interest, co-analysing study results, advising if researchers' ideas and thoughts are on the right track (e.g. appropriate wording used or right engagement strategy put in place), and public contributors doing sense-checking and contributing to potential engagement strategies with the broader public. The following quote shows the variety of involvement and its impact:

> "*Extremely impactful, (…), it's actually led to changes in the direction of our work, but in cases where that hasn't necessarily happened, that they've been more supportive of what we're kind of thinking and it has changed the way that some are kind of like thinking about the topic of public trust and public confidence, for example, and we only ever used to think like the wording that we would use as an organisation was we need to earn public trust. We need to build public trust but then through the [advisory board] through exploring that a bit more, we've kind of changed our way of thinking, so it's more about demonstrating trustworthiness in the use of data and building public confidence*." (Harriet)

This theme has shown that public involvement should not be an afterthought and needs to be a genuine (but often time- and resource-consuming) process that can have a significant impact on researchers' work. This can be especially seen in the following extract:

> "*It is difficult to do really well, and it takes a lot of time and a lot of resources, and I think people underestimate that. I also think there's a culture towards PPI as a tick box*." (Penelope)

10.7. Discussion

1. Provide information in lay language and, where not possible, explain in simple English. Ensure these explanations are available at any point to the public contributor (e.g. through an online dictionary).
2. Rotate public contributors on a “big data panel” every 3 years to bring in new ideas and lay perspectives.
3. Reach out to new communities for at least 50% of the new attendees, potentially using charitable/partner organisations to help.
4. Identify relevant seldom-heard communities for each project.
5. Consider strategies to add additional diversity on multiple characteristics (e.g. LGBTQ+ and ethnic minority, or disability).
6. Adequate and ongoing training/support for PAs should be provided to empower them so they can truly contribute.

Table 11 Key recommendations on involving and engaging seldom-heard communities around big data research

Our findings have shown that talking about big data ‘with’ (rather than ‘to’) public contributors can be challenging, but that PPIE can be meaningful for both researchers and public contributors. The findings elicited how researchers, and their research can benefit from involving and engaging seldom-heard communities. Table 11 summarises key recommendations. This adds to the previous literature on meaningfully including a diverse range of communities (Taylor *et al.*, 2018) and is relevant to other areas of health and social care research. PPIE requires time and resources (Boden *et al.*, 2021) and not all communities are often equally involved (Beresford, 2013). However, our participants have shown that inclusion around big data research (because of the complexity of the topic) takes additional time and resources to succeed (even in contrast to other heath research). This can be seen in extra activities such as a “pre-engagment engagment" which was suggested as a baseline for successful working with the community. Our findings challenge the perspectives of some researchers who believe that public contributors rarely care about or can understand big data research and thus are not able to be involved in decisions around whether medical data can be reused for research (Mouton Dorey, Baumann and Biller-Andorno, 2018). Involving and engaging seldom-heard communities in big data might be more challenging than in other forms of health research but it is important as big data research offers an opportunity to reduce health disparities (Zhang *et al.*, 2017). Without seldom-heard voice input this might not happen.

The findings confirm that defining a group as a seldom-heard group is context-specific (NIHR, 2020; Paprica *et al.*, 2020). The participants named numerous types of seldom-heard communities involved and engaged within the context of their work. Researchers should reflect on who would be the most seldom-heard group within the context of their study and recognise that this might include more than one community. The concept of superdiversity (Bradby *et al.*, 2017; Kirwan, 2021) could provide researchers with further guidance on moving away from looking at single characteristic (e.g. ethnicity) of the community and focusing instead on diversity within diversity. This would ensure that the needs of communities within communities are considered.

Researchers need to take time to plan PPIE well as they design their projects. NIHR guidance (2020; 2021c) recognises this and recommends working with communities on a long-term basis. Our findings have shown the importance of building and maintaining relationships with organisations, especially charities. This confirms previous research that shows links to the third sector are crucial in building trust (Steel, 2005a; Morrow, Boaz and Sally, 2011; Nelson and Burns, 2020). They often act as gatekeepers but also have the potential to act as a partner. There is, however, a risk that researchers would not reach many communities as they might be limited to the partner organisation’s level of diversity.

There is a growing trend to establish a pool of volunteers interested in participating in PPI activities (Grotz, Ledgard and Poland, 2020). This approach might appeal to those who have time, resources and feel comfortable with working with institutions. However, this risks public contributors becoming ‘usual suspects’ of people who are involved regularly and thus not providing new contributions. There is the danger that they will become more expert than researchers themselves; thus, no longer providing lay experiences and views in the project. There remains a contentious issue, how to strike a balance between public contributors being capable of contributing fully but also retaining a lay perspective (Ives, Damery and Redwod, 2013). One of our participants suggested the need to change public contributors on advisory boards every three years. This offers a solution to deal with the challenge of ‘usual suspects’ and brings a fresh public perspective but adds more work on the part of the researchers to recruit, provide training and support new public contributors on the project. The other option is to sense-check any work with the broader public.

Researchers should also ensure that any involvement is not tokenistic and enables power-sharing between researchers and the public contributors (Ocloo and Matthews, 2016). There is no one ‘right’ way to do it, and the approach depends on the project’s needs (or resources) and the public contributors' interests. However, their interests should not be confused with their understanding of the topic, and researchers should provide training to improve public contributors’ knowledge, thus facilitating their ability to contribute. This genuine empowerment was seen as crucial among our participants when discussing big data research with public contributors. Although not mentioned by our participants, some public contributors, for example, coming from Indigenous communities, might also require researchers to respect their values to feel truly empowered (Rowe *et al.*, 2021).

10.8. Study limitations

The study participants came from diverse communities, e.g. various ethnic minority backgrounds. However, we did not record if they are a part of other seldom-heard communities, e.g. LGBTQ+ or people living with disabilities. We only explored the perspectives of the researchers, and there is a possibility that the public contributors (including those coming from seldom-heard communities) would have a different view on their PPIE activities around big data research. As big data is a fast developing and diverse research area, new ways of involving and engaging will emerge, so future research should further explore how researchers involve and engage public contributors and how concepts of super diversity could be utilised.

10.9. Conclusion

Our study explored how researchers involve and engage public contributors (especially seldom-heard communities) in a meaningful way in big data research. The findings highlight that there is no one right approach to doing PPIE and that PPIE strategies are project specific and depend on the public contributors, researchers' needs, resources and time available. We encourage others to reflect on their involvment strategies and hope these results will support researchers who want to involve more seldom-heard communities in complex research topics such as big data.

11. Polish and South Asian communities' study

This chapter presents the findings from the third primary study as published verbatim in the International Journal of Population Data Science titled *Exploring how to improve the involvement of Polish and South Asian communities around big data research. A qualitative study using COM-B model* (hereafter "Polish and South Asian communities' study) (Teodorowski *et al.*, 2023b). This research builds on previous researchers' study by exploring the involvement from the perspective of members of the public. Before I present the paper, I explain how this study links to the literature review. After that, I discuss how the study's aims and the findings add to the existing knowledge.

11.1. Linking the literature review with the paper

The relevant issues from the literature review for the Polish and South Asian communities' study were:

- Literature on how to involve seldom-heard communities in big data research from the perspective of members of the public is limited.
- There is no guidance for researchers on how to involve seldom-heard communities and what factors could influence people's involvement.
- Involving people in big data research can include additional barriers, but further challenges could be experienced by public contributors who come from seldom-heard communities. However, there is limited understanding of how this would apply in big data research.
- Seldom-heard communities, especially ethnic minorities, might have lower trust than the general public that big data researchers would use their data responsibly.

This study addressed these gaps in the literature by exploring how Polish and South Asian communities in the UK can be encouraged to be involved in big data research. Second, the study explored the participants' views towards big data research. The approach to address these research aims was to conduct semi-structured interviews with twenty participants in each ethnic group who were living in the UK but had not been previously involved as public contributors at the time of the study. To further understand what could influence members of the public to become involved as public contributors in big data research, the study utilised the capability-opportunity-motivation-behaviour (COM-B) model. The study findings were mapped under three constructs of the COM-B model to explore multiple factors influencing involvement.

The study offers novel findings around involvement in big data research from the perspective of Polish and South Asian communities. The findings indicate that multiple factors could influence members' of the public decision to become public contributors. The COM-B model assisted in identifying and exploring the relationship between them. However, the study suggests that researchers who wish to involve Polish and South Asian communities in their work might not be able to influence all factors that impact members' of the public decision to be involved in big data research. The study was the first one using the COM-B model to understand what could influence people's decisions to become public contributors. It suggests the model could benefit other researchers when developing their involvement plans. In the discussion chapter of this thesis, I discuss further the COM-B model and provide guidance on how to apply it in other public involvement initiatives.[58] Last, findings offer new insights into Polish and South Asian perspectives towards reusing their health data for research.

11.2. Abstract

Introduction

Involving public contributors helps researchers to ensure that public views are taken into consideration when designing and planning research, so that it is person-centred and relevant to the public. This paper will consider public involvement in big data research. Inclusion of different communities is needed to ensure everyone's voice is heard. However, there remains limited evidence on how to improve the involvement of seldom-heard communities in big data research.

Objectives

This study aims to understand how South Asians and Polish communities in the UK can be encouraged to participate in public involvement initiatives in big data research.

Methods

Forty interviews were conducted with Polish (n=20) and South Asian (n=20) participants on Zoom. The participants were living in the United Kingdom and had not previously been involved as public contributors. Transcribed interviews were analysed using reflexive thematic analysis.

Results

We identified eight themes. The 'happy to reuse data' theme sets the scene by exploring our participants' views towards big data research and under what circumstances they thought that data could be used. The remaining themes were mapped under the capability-opportunity-motivation-behaviour (COM-B) model, as developed by Michie and colleagues. This allowed

---

[58] See more at page 226.

us to discuss multiple factors that could influence people’s willingness to become public contributors.

Conclusions

Our study is the first to explore how to improve the involvement and engagement of seldom-heard communities in big data research using the COM-B model. The results have the potential to support researchers who want to identify what can influence members of the public to be involved. By using the COM-B model, it is possible to determine what measures could be implemented to better engage these communities.

11.3. Introduction

Involving the public in research design and execution has become firmly embedded within the UK academic research environment (NIHR, 2019). Public involvement has had a positive impact on research (Brett *et al.*, 2014) and moral and ethical principles support public involvement (Thompson *et al.*, 2009). It offers the researcher a lay perspective, especially from patients who have a lived experience of the health conditions being studied (Rhodes *et al.*, 2002) and has the potential to improve the quality of research. In this paper we focus on public involvement in big data research (Aitken *et al.*, 2019). There remain disagreements about what constitutes big data (also called data-intensive research (Mehta and Pandit, 2018) or routinely collected data (Newburn *et al.*, 2020); we define it as linking data from different sources (e.g. routinely collected patient records, genomic data etc.). These data are often reused for a different purpose from the initial direct care purpose for which they were collected; a secondary research purpose.

There are various ways of describing public involvement in research. We follow the National Institute for Health and Care Research (NIHR) definition of involvement as “research being carried out ‘with’ or ‘by’ members of the public rather than ‘to’, ‘about’ or ‘for’ them” (NIHR, 2021a). This differs from public engagement and participation. Engagement is about sharing information and knowledge about research. Participation refers to study participants such as interviewees or those enrolled in a trial. Throughout the paper, the term public contributors will be used to designate those who are involved in research as lay people, who are not researchers or healthcare professionals, and provide the public voice on research projects.

Public involvement can provide legitimacy for research (Manafo *et al.*, 2018). The public might feel disconnected from research (Aitken *et al.*, 2019) and public involvement can bridge the gap between big data researchers and members of the public. This is especially important because it could be difficult to reuse health data without the public’s support. A systematic review by Aitken and colleagues has shown that the public generally support big data initiatives but still have concerns about how their data are being reused (Aitken *et al.*, 2016). Ensuring social licence for research, which is more than meeting legal requirements requires trustworthiness and addressing public concerns is important (Jones *et al.*, 2020a; Muller *et al.*, 2021). Public contributors have been previously successfully involved in big data projects. One such initiative that uses linked health data for research is the SAIL Databank, that has a public panel. Their role is to advise on the work of the Databank and the proposed use of data (Jones *et al.*, 2020b; Jones *et al.*, 2019; Jones, McNerney and Ford, 2014). Public contributors can also take part in public deliberations on linking data for research or its management (Teng *et al.*, 2019; Deverka *et al.*, 2019), co-design ways of engaging with the broader public (Townson *et al.*, 2020) and be a public representative (contributor) on the project's advisory group (Newburn *et al.*, 2020).

All sections of society need to be included in public involvement activities; the onus is on researchers to ensure equal opportunities and that potential barriers to involvement are overcome (NIHR, 2021a). Not including a range of communities can increase health inequalities if the views of some groups are missed. Especially those seldom-heard groups, who often experiencing health inequalities (Hogg, 2009). Researchers and policymakers must include diverse voices to ensure that services can effectively respond to all communities' needs (Phillimore, 2011). Seldom-heard communities are a broad definition, so we decided to focus on migrants and ethnic minorities as these are recognised as under-represented communities (e.g. NIHR (2020)). Public involvement can help in formulating initiatives to reach seldom-heard communities (Nelson and Burns, 2020).

Despite recognising the importance of these kinds of initiatives, there is limited literature on how to successfully involve seldom-heard communities in health and social care research (Ocloo and Matthews, 2016; IPFCC, 2018), and there is a particular lack of literature on how to involve people in big data research. A bibliometric review by Boote and colleagues (Boote, Wong and Booth, 2015) of public involvement in health research was conducted between 1995 to 2009. These results are now relatively old in this quickly advancing field. Still, they show that there were fewer examples of involving public contributors in quantitative research than in qualitative research in that time period. This could be explained by public contributors finding it easier to understand qualitative data such as interviews or focus groups rather than interpreting results from statistical models (Boote, Wong and Booth, 2015). This could be even more of an issue in big data research because one of the challenges might be explaining data linkage and processing concepts in lay language (Townson *et al.*, 2020; Deverka *et al.*, 2019), and training on technical terms might be required (Jewell *et al.*, 2019).

People’s attitudes towards being involved in research could be influenced by personal and social motivations (Turk *et al.*, 2020; Tarpey, 2006), including their available time. Additionally, involvement requires additional time and resources from the researcher (Boden *et al.*, 2021). However, our understanding of why people get involved is limited. According to our knowledge, no studies have explored public involvement in research with people who have not been previously involved. Therefore, there is a need to address this gap in the literature and explore what influences people to become involved in big data research as public contributors.

11.3.1. Theoretical underpinning

There has been a shift in the public involvement literature from blaming individuals (around members of the public not becoming involved) to collective responsibility (often moving to this becoming the responsibility of researchers and research environments) for members of the public not being involved (NIHR, 2020). This can be seen in changes in the wording used by researchers. These communities are now described as seldom-heard or under-served rather than hard-to-reach groups. The latter places the responsibility on the public, whereas the former the responsibility is on researchers (NIHR, 2020). For researchers to know what they can do to involve seldom-heard communities more successfully, a theory is needed to understand what influences members of the public to become involved in research. A wide range of behaviour models can be used to identify the ideal conditions for a behaviour change intervention to be effective. We adopt the well-established capability-opportunity-motivation-behaviour (COM-B) model created by Michie and colleagues (Michie, Van Stralen and West, 2011). The COM-B identifies three factors that need to be present for any behaviour to occur: capability, opportunity and motivation. We consider public involvement as a particular behaviour that can be influenced by factors that can be categorised under these headings. Each of them is equal in value and contributes to behaviour change (Michie, Van Stralen and West, 2011). However, it does not mean that researchers would need to equally influence each of them, even if they can.

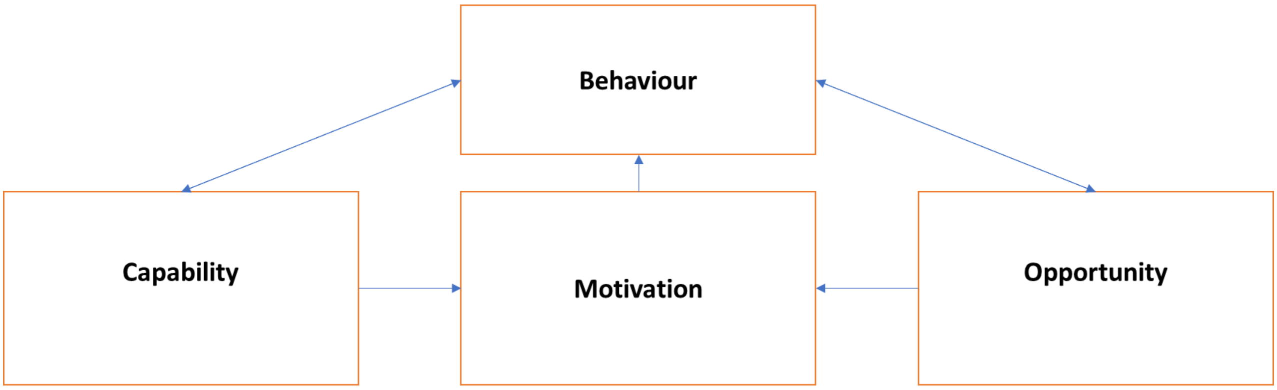


Figure 9 COM-B model as developed by Michie and colleagues

We adopted the COM-B model (Figure 9) in our study as we recognise that it is likely that multiple factors could influence changes in peoples' willingness to become public contributors. Secondly, using theories in understanding behaviour can be more effective than interventions without theoretical underpinning (Michie *et al.*, 2014). As far as we know, this is the first time the COM-B model has been applied in a study exploring public involvement. Previously, it has been successfully utilised to develop effective healthcare interventions (Barker, Atkins and De Lusignan, 2016; Alexander, Brijnath and Mazza, 2014) and to understand health behaviour change (West *et al.*, 2020; Rubinstein *et al.*, 2015; Willmott, Pang and Rundle-Thiele, 2021).

11.3.2. Research aims

This study aims to understand how we can increase the inclusion of South Asians and Polish communities in the UK in public involvement initiatives in big data research.

Our study aim was achieved by answering the following research objectives:

1. Exploring South Asian and Polish peoples' views on how big data are collected and used.
2. Understanding enablers and barriers in involving South Asian and Polish communities in public involvement initiatives in big data research.
3. Exploring how researchers could involve members of the public in the design and governance of big data research.

11.4. Methods

We used the Standards for Reporting Qualitative Research (O'Brien *et al.*, 2014) to ensure transparency when reporting this research. The completed guidance is available in appendix 10.

11.4.1. Participants

Our participants are from Polish and South Asian communities in the UK. People from Poland are the second largest migrant group (based on country of birth) and the largest (based on nationality) in the UK (Vargas-Silva and Rienzo, 2022). They used to be the largest migrant community (based on the country of birth) in the UK until 2018 (Vargas-Silva and Rienzo, 2022). This decline is associated with the UK's decision to leave the European Union, which removed the legal right to migrate to the UK for those with a Polish passport. Those who already were living in the UK had to apply for settled status and their right to remain. Many Polish people living in the UK experienced racism and discrimination (Rzepnikowska, 2018) and Brexit as a political event negatively influenced Poles' mental health and wellbeing (Teodorowski *et al.*, 2021b).

South Asians consist of multiple national groups: India, Pakistan, Bangladesh, Nepal, Bhutan, the Maldives and Sri Lanka. Despite not representing one nationality (as the Polish do) and coming from various cultural backgrounds, they are often grouped for health research under the category of South Asians as they share similar health experiences (Fischbacher, Hunt and Alexander, 2004; Ranasinghe *et al.*, 2013; Kakde, Bhopal and Jones, 2012). Based on the country of birth, Indians are the largest and Pakistani are the third largest migrant community in the UK. However, based on nationality, Indians are the third largest and Pakistani the eighth. South Asian communities offer another angle of being a migrant or ethnic minority community in the UK. South Asians experienced British colonialism and hold different migration status in recent years and they also experience racism (Prajapati and Liebling, 2022).

For clarity, we refer to both Polish and South Asians as participants. The sampling was purposive to select participants that enabled us to address the research aims. Participants had to live in the UK, identify themselves as part of Polish or South Asian communities (but not necessarily have that nationality) and to never been involved as public contributors. The last requirement ensured that we could understand the views of members of the public who had not yet been involved in research as public contributors. All participants were over 18 years old. Forty people took part in interviews, twenty Polish and twenty South Asian participants. The South Asian participants came from the following communities: India (n=11), Pakistan (n=4), Nepalese (n=3), Bangladesh (n=1) and mixed Asian (n=1). We did not collect other participant demographic characteristics.

11.4.2. Procedure

Recruitment was through social media (Facebook, Twitter and WhatsApp). We promoted the study with existing groups and our own networks; we did not use paid adverts. Interviews took place between October 2021 and March 2022. Potential participants contacted the researcher to discuss their participation in the study and received a participant information sheet before deciding if they wanted to participate. Prior to the interview, the authors sent them a consent form to be signed and returned before the start of the interview. PT conducted interviews on

Zoom to ensure that data could be collected safely during the Covid-19 pandemic (Lobe, Morgan and Hoffman, 2020). Potential Polish participants had the option of receiving a participant information sheet and consent form in Polish or English. This was translated by the Polish-speaking author (PT). South Asian communities use multiple languages, and we did not have translation expertise in-house, thus providing information in other languages was not feasible. However, South Asian participants were offered interpreters to support their participation. The majority of interviews were conducted in English, some in Polish with a native Polish speaker (PT) and one with the support of an interpreter (in Urdu). We used a professional interpreter, who received information about the study in advance and attended briefings with the researcher before and after the interviews. This allowed us to record any cultural or unsaid messages as reported by the interpreter. A semi-structured interview guide steered the interview (see appendix 2). All participants were offered a £20 shopping voucher as a thank you for their time.

When conducting an interview in one language but presenting the findings in another, there is a danger that the essence of the message shared by a participant will be unintentionally lost (Squires, 2008). We followed Gawlewicz (2016) guidance to overcome this challenge and noted any potential discrepancies. The process included creating notes when there were unspoken assumptions (e.g. mentioning ‘church’ by Polish participants would refer to the ‘catholic church’), the response was culturally influenced, or hard-to-translate words were used.

All interviews were transcribed and anonymised. We use pseudonyms when reporting the study findings.

11.4.3. Public contributors

Involving public contributors in the qualitative analysis can improve the quality of findings (Garfield *et al.*, 2016) Two public contributors (SA and NT) were involved in all stages of the research project: design, recruitment, analysis and the write-up. Public contributors were recruited through NIHR Applied Research Collaboration North West Coast where they receive ongoing support and training. Therefore, they were not participants in this study. They received reimbursement for their involvement. They are both experienced contributors and have been involved throughout the doctoral research of PT. SA is Data Ambassador for Care and Health Informatics theme within the NIHR Applied Research Collaboration North West Coast. At the design stage, public contributors were involved in creating the interview guide. As a part of it, we included a short description what big data research is with the aim of providing an explanation to participants and we wanted to ensure that this description was written in lay language. Hence, the public contributors and PT organised two sessions with members of the public (n=9) who were not previously involved in big data research to pilot the topic guide. These members of the public were not later participants in the study.

PT delivered training on reflexive thematic analysis for public contributors. This included providing information about the process and practical exercises to build public contributors' confidence in qualitative data analysis. Involving public contributors throughout all stages of the analysis provided a lay perspective and understanding of participants' perspectives (Hemming *et al.*, 2021). For example, they were able to point out nuances in the initial interview that led to additional questions for the following interviews.

11.4.4. Data analysis

All interviews were uploaded to NVivo 12 and analysed using reflexive thematic analysis (Braun and Clarke, 2006; Braun and Clarke, 2021c; Braun and Clarke, 2019). Two public contributors (SA and NT) jointly coded with PT one interview from each group of participants (Polish and South Asians participants). Later PT coded the remaining interviews. Data saturation was reached when no new codes or themes were identified (Guest, Bunce and Johnson, 2006). We met as a whole team to discuss and reshape themes. Our research team also provided both insider and outsider perspectives on the studied communities. PT comes from the Polish community and SA and NT are from a South Asian background. The initial analysis was inductive, and then categorised into the components of the COM-B model.

11.4.5. Ethics

Institute of Population Health Research Ethics Committee at the University of Liverpool approved the study (number 10063).

11.5. Results

In our analysis, we identified eight themes. First, we set the scene by discussing a theme called 'happy to reuse data'' that shows our participants views towards big data research. We present them mapped under the COM-B model, the factors of capability, motivation and opportunity (Figure 10). The themes appeared across both participant groups, indicating the results could be applicable to other migrant and ethnic minority groups in the UK.

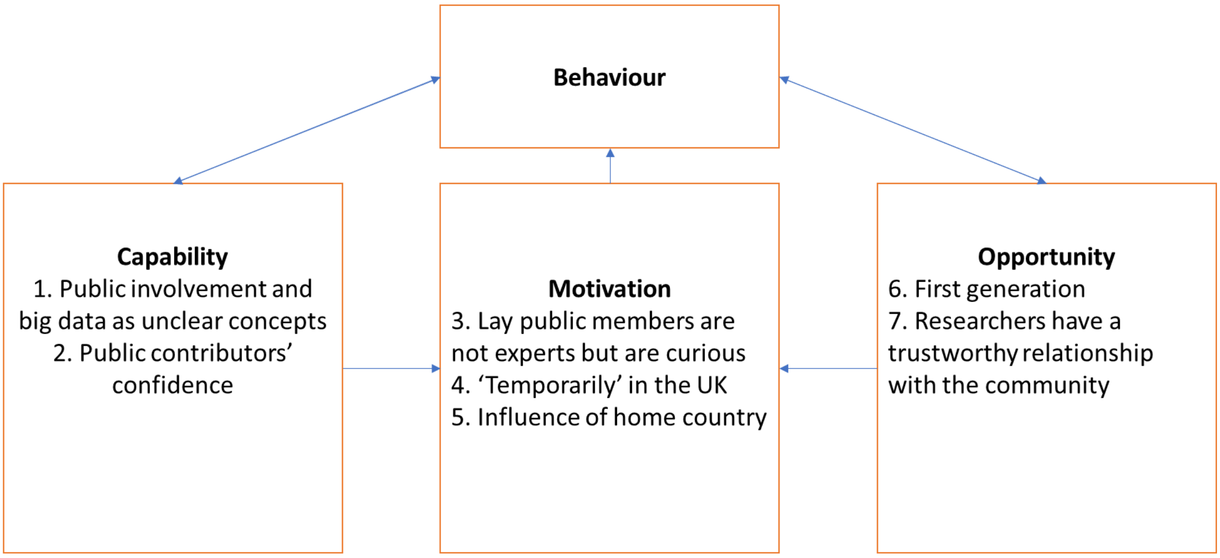


Figure 10 Themes mapped under COM-B model

11.5.1. Happy to reuse data

This theme elicits participants' perspectives on big data research. The majority of them supported the reuse of medical data for research with a belief that access to their medical data are needed for science to progress. They recognised that this new type of research drives changes and can help develop new medical treatments and interventions. However, they want their data to be used only for a 'good purpose'.

> "*Good purpose, so in the future can be used to help people somehow. So if this is medical research which will help us to discover new medication, for example, or new ways of treatment, why not? I don't see the reason why not to share my information with others*." (Ada, Polish)

The issue of what constitutes a 'good purpose' and under what circumstances research can take place was explored further with the participants. Participants felt comfortable with their data being securely processed. This should consist of the following characteristics: secure, anonymised and responsible storage, authorised access, and public ownership. We will discuss each of them.

Participants felt that their data should be stored securely. However, the majority admitted that they were not experts (due to this being a technical topic) on what this should look like. They pointed out that they are happy to share any data as long as it is anonymised. This can also include sending data abroad for research, but here some participants had concerns that other countries might not follow the same data protection legislation.

> "*I don't know somewhere, somewhere safe. [laughing] Yeah, but you know we talk a lot about you know… There's different type of encryption, etc, and so I'm not really like the you know much. You know very much technical to know all the details, but it definitely should be encrypted… data on some safe servers I guess*." (Zosia, Polish)

When discussing if private companies (e.g. pharmaceutical companies) should have access to medical data for research, participants had mixed feelings, some of them felt that only universities should be able to access medical data for research. Participants described universities as public bodies whose work everyone can benefit from, whereas private companies exist for profit. Others recognised that private medical companies might also require it for research (e.g. evaluate if their new medicine works). However, some found it unappealing and recommended that, in this case, additional consent might be required. Joint projects with universities made private companies' access to data more acceptable for the participants. This extract shows how most of the participants who were against private companies having access to their medical data for research answered when it was suggested that it would be a joint project between a university and the private company:

> "*That will be OK if they work with like university researchers that that will be perfect (…) if it's sort of like a research approved by university researchers, that's fine*." (Irena, Polish)

Participants felt that any data access needs to be authorised. Two levels of consent were thought to be desirable among the participants. First, on the individual (patient) level and second, organisational (data access) level. If the public truly owns their data, participants want to consent themselves to their data being reused. Further, consent should be related to individual projects. Participants did not always perceive opt-out consent in positive terms. The reason behind this view was influenced by their preference to be informed how their data is used. Others felt that they might be willing to support one project but not the other. Secondly, at an organisational level, authorisation is needed by a governing body that gives permission for researchers to access data for research. Healthcare professionals were often named as those who should participate in the decision process. Participants felt that public contributors should be involved too. However, it remained unclear what the background of these public contributors should be. Some raised concerns that any (public) contributors should be experts and have a background in big data. This ensures that they understand the topic of discussion and can provide useful feedback.

> "*There should be common public there, because the voice of the people you know. It's important that people are heard but those people I think it's important that they need to have some medical background, so there should be people either studying medicine or*

> *somehow related to medicine somehow related to clinical research. You know, and then it will be beneficial. Otherwise, I don't think if you just put a layman sitting in the community and you're talking about how the blood samples will be taken. Then some tests will be done. I don't think a layman will understand much*." (Jai, South Asian)

Even when routinely collected health data are reused for research, participants believed that they are the primary owners of it, as it relates to their own medical conditions, life and experiences.

> "*I would be the primary owner of my data*." (Yashica, South Asian)

This view affects how participants perceive other ethical issues around big data research, especially consent. People should decide if their data are reused, and some participants also recognised the need to have access to their data themselves or be aware if it is reused. This information should be available in a transparent way. The following quote emphasises the issue of explicit consent:

> "*So we should own the data as we own our properties and unless I will give explicit consent to use my data, then no one can use it. So let's say I have an NHS app and and then I have my medical records and my history of my GP visits and vaccination. If I'm not giving explicit consent. It shouldn't be used. The data shouldn't be used. And every time, obviously that goes, you know you've got different, even the implication of that, because I who should be allowed to withdraw at any time. You know that you know, you know, the ethical consideration I should be allowed to access data anytime. What did you see? What did you do with it? And things like that*." (Greta, Polish)

This theme has shown what participants perceive as ‘good’ or appropriate reuse of their medical data for research.

11.5.2. Capability

11.5.2.1. Public involvement and big data as unclear concepts

The majority of participants had a limited understanding of public involvement and/or big data research. Those who were interested in technology (e.g. through work) or health research expressed some knowledge of big data research. The following quote illustrates how surprised many participants were when asked how they felt about routinely collected data being reused for research purposes:

> “*I never thought about it, you know whether they would be using my data for any research, but after reading it and after your question, I'm thinking they might have used in to see, like you know when they were providing a vaccination or for COVID or*

> *anything like that, you know they might use my data. But really speaking, I never thought how they're using it. I thought it is only private to me."* (Bhavna, South Asian)

They often felt more comfortable referring to non-health-related examples. These were usually negative and thus could influence the public's views on how their data are (and if should be) reused.

> "*We all know that how Facebook then Google use your data and then how it affects your searches and the advertisements you see. And how it bombards you to their benefit. You know with the things that you see is not actually what you want to see just because you have spoken about. Let's say you want to buy a sofa. You'll only see the sofa advertisements. And if you want to buy a table, you'll just see the table advertisements. Because people are tracking you. So I think people are sceptical about the fact that, they don't want their data to be shared anymore. Whether it is mobile data or it is any information. So I think that is the biggest barrier that the invasion into privacy due to Facebook and Google and you know Twitter and Instagram. I think that invasion has caused people to be sceptical about everything." (Jai, South Asian)*

Participants agreed that there is a need to raise awareness about big data research and public involvement as they felt that the public do not know about opportunities or what kind of research is being conducted. Without such awareness, they would not be able to get involved. All information about opportunities should be provided in lay language. There should be one point of contact to learn more about big data research; some recommended their GP to be the first place to learn more about it. However, this raises questions about the practicality of these expectations.

11.5.2.2. Public contributors' confidence

This theme links to the previous one, because participants with limited knowledge of big data research also had lower confidence in becoming pubic contributors. However, that confidence improved when language barriers were overcome. The majority of those participants whose interview was conducted in their mother tongue felt that the language barrier could influence how much they can (and are willing) to be involved. Their limited knowledge of English influenced their confidence levels, especially when the involvement was related to such complicated topics (as they perceived big data research to be). The following extract shows that even those who spoke in English were concerned about their language skills:

> "*We think before we know how will be answered to be very frank with you and to have a call with you. I was thinking you know how will be my English. You know whether I'll be comfortable whether I can understand you, whether I can answer your questions, whether*

> *you know, such things that will be always running in the background because we're not confident.*" (Bhavna, South Asian)

Participants recognised that there is a need for willingness for both researchers and the public to understand each other; many considered that learning English is a must as they live in an English-speaking country. However, building confidence and fluency takes time. They also predicted that many newcomers would not do it. However, researchers could support those with limited knowledge of English to become involved as public contributors. Participants suggested translated materials, the presence of interpreters (both professional as well as from the community) or sessions in their mother tongue (e.g. hiring a researcher speaking their mouth tongue). Having an interview in their mother tongue would make people more comfortable with expressing their ideas. Researchers coming from ethnic minorities could help public members to overcome this language barrier. The following quote shows how a participant described the Polish-speaking researcher during the interview:

> *"You [referring to interviewer] are this kind of bridge, a Pole who knows Polish and studies at a British university, so you are a bridge that could just be offered to Poles who do not speak English well and do not understand English well." (Mateusz, Polish)*

The choice of language in which information is presented is crucial. As one of the participants who has experience of sharing information on social media pointed out, presenting the information initially in the target audience language can have higher visibility than in English.

> *"All articles [on social media] are in two languages in English and Polish, but if it starts with English, not Polish, it has 85% have smaller ranges. When it starts with Polish, it grows significantly. (…) Now imagine yourself, you scroll with your thumb down. And you only see the first two, three sentences (…) you see them in English. Well, then you are skipping it."* (Filip, Polish)

This theme has shown that participants' faith in their English language abilities influences their desire to become public contributors. Researchers should provide support for those who have limited confidence in this regard to allow seldom-heard communities to get more involved.

11.5.3. Motivation

11.5.3.1. Lay members are not experts but are curious

This theme elicits participants' perception that they are not experts on this topic, so they are unsure if they should or could be public contributors. Lived experience is a key part of being a public contributor in research, but our participants did not see themselves in that way, or think that their personal experiences could contribute anything meaningful. At the same time, they were curious to learn more about big data and public involvement. Some participants felt that when discussing complex issues, researchers know better, and this quote shows this challenge:

> "*I think it circles back to that one point, say, as long as those people are of that particular field it, it would be really helpful to get their advice. Say for example, say, you are doing some research in urology and there is a renowned doctor or staff like that in urology. And if you ask for his advice, that's good. That's a good advice which will benefit both the research group and it's like processes and its policies. But if you ask me. Me being an engineer having no concept of anatomy or anything else. So my advice might derail the entire research project. It could be like some generic if you ask me some generic policies, yeah, but if you are like specifically asking domain specific policies or domain specific processes then, no, I wouldn't be like comfortable to be an advisor*" (Onkar, Polish)

Not everyone was interested in becoming a public contributor. Participants were curious and wanted to learn more, but this did not mean that everyone would be interested in getting involved or involved to the same extent, so researchers need to reach the right people and listen more to the public on what they expect from such involvement. Thus, public involvement is more about quality than quantity. The participants spoke about different ways of reaching the community, recognising that there is not one way to reach everyone. These methods included directly reaching the community through ethnic minority media, translated leaflets at shops with products from their home countries, and reaching places of worship (mosque or church) and ethnic minority organisations.

> "*There was a Scottish and Polish priest [as Poland is a majority catholic country, the participant means catholic priest] (…) if we would like to say something to the Poles, or one could ask to appear and say something at the end of the mass, there is no problem with that. It is already such a larger group of people, because this is a mass for Poles at a given hour, so it is known that most of the people at the mass are Poles, so it is also such a place where you can convey something to who is interested in some study*." (Agata, Polish)

Some participants pointed out that among their communities there remain some even more seldom-heard groups. They saw them as people who struggled to find activities for themselves in the UK. An example would be the older or women. Many Polish people bring their parents to the UK for family reunions but they often do not speak the language and do not know anyone, so they struggle to socialise. The next extract shows how they organised a local

Polish group for older people that later was suggested as a place for researchers to use to offer activities.

> "*We created a Polish Senior Club, because we noticed, of course, in connection with covid, that the elderly are simply lonely and we just wanted to change something about it, because children still have some attractions, and so on. Well, unfortunately ... But unfortunately the elderly are forgotten [...] some of them did not know it at all Poles of their age also the club is 50/55 and above*." (Dagmara, Polish translation)

All these ways of reaching the community and those doing it need to be seen to be trustworthy, and the participants called for more active communication between the public and researchers. They want researchers to listen to their feedback and have feedback back about their contribution.

> "*I just like to know I'd like to hear from, I'd like to see what other people think about this and just suppose kind of learn from their opinion, learn from them to see and then (…) it would be good to know like the what's going to be done with the with my input as well. OK, we're going to. I'd like to, you know, keep up to it… Yes, would like to keep up to date. Keep updated with the progress of the project as well*." (Pourang, South Asian)

This theme has shown that researchers need to shift the public's perspective on how the public can contribute to big data research. People are curious, and there are ways for researchers to reach seldom-heard groups successfully.

11.5.3.2. 'Temporarily' in the UK

This theme focuses on participants' perspectives on how their integration and length of residency in the UK could influence their willingness to become public contributors. Some participants reported that they are only here on a temporary basis. This could be due to their circumstances (e.g. coming to study) or visa status (e.g. temporary right to remain). Because of this, they had not settled in the country (yet, or they do not plan to). Thus, they felt that they should not be involved in their community or as public contributors. If they are living in the UK temporarily, then they should not be involved in decision-making at this level. However, this issue was raised by participants who have lived ten years or more in the UK.

> "*Well, we are in a foreign country, we are not in Poland, we are not citizens ... I mean a citizen. We are immigrants. [laughing] We are not British citizens, we have these statuses [referring to settled status that was given to the European Union citizens post-Brexit]. Yes, we have the same rights as the British, but really, but despite everything I do not know*

> *how and I know many people too, always have the back of the head that we are not British and that it is not quite like that it is our home, not our country*." (Marlena, translated from Polish)

Some participants felt that they would not benefit from their involvement because they are only in the UK for a short time. This was visible with some South Asian participants who only had temporary visas and if they wished to stay longer, would have to renew them.

> "*No, I think the barrier can only be that, you know, although you assume that people are working here, sometimes people are only here. You don't know their residential status, that they might only be here for a contract job, or they don't intend to stay here very long so they don't feel they need to take part in any of such research because they are, they might only be living in this country for four years or five years and then going back. So that could be another barrier as well that you don't know sometimes who you're targeting*." (Jai, South Asian)

Despite migration status differences (e.g. different migration rules governing the right to remain) between Polish and South Asian participants, the temporality issue was shared among both groups. Researchers could have a limited scope to influence this factor as this depended on personal circumstances, current legal immigration and right to work rules.

11.5.3.3. Influence of home country

The last theme relates to motivation and is about how the experience of living in one's home country could influence becoming a public contributor. Some participants recognised that a Polish or South Asian mentality could influence how (and if) people get involved in big data research. For example, the experience of communism among Polish participants influenced how they perceive governmental or public institutions; volunteering might differ in the UK from the types they had experienced before or there is no equivalent for the role of public contributor in their home country. The following quote shows how such views could influence involvement:

> *"Well, we are such a society that it is difficult for us to please, or even if you try something, we always see some negatives, and we do not want to do this. We can complain rather than to take part in something, to help each other tell each other what does not fit (…) We are such a strange community that we do not want to, we complain about the health service, we complain about everything, about politics, about Tusk [former Polish prime minister and one of the opposition leaders, often illustrated by the current government as a symbol of everything that went wrong], and we don't really change anything"* (Inga, Polish)

Although it would be challenging for researchers to approach all the issues around peoples' existing views on public involvement, it is possible to engage in a two-way conversation to discuss what opportunities are available in the UK. This would require researchers to better understand these groups perspectives on key issues.

11.5.4. Opportunity

11.5.4.1. First generation

The theme shows that participants face barriers such as lack of time, childcare, and travelling costs, for example, to become public contributors. These barriers could be experienced by anyone who wants to get involved in research. However, participants felt that ethnic minorities (especially migrants) often do not have access to support networks (e.g. family) because they kept moving around the UK or did not fully feel integrated into the community. That experience was described as the result of being 'the first generation in the UK' and explained by this participant:

> "*It depends on how much time we have spare on doing such kind of thing. Because come here, I would say I am the first generation that is being in UK, so the second generation is just out two or three years now. Whereas when (…) I think about my parents. They always had help around, you know they had their granny, granddad to take care and to have different things and time management apart. I could say if I now don't have anybody to take care of my kid. I should rely on a daycare or this school, or when they're in school that is the only time I get*." (Ridhi, South Asian)

As the interviews took place during the Covid-19 pandemic, the impact of this was seen as an additional factor. Participants spoke about how Covid has changed the way people communicate. Many participants spoke about the benefits of using social media in reaching and working with communities. However, some participants still preferred face-to-face meetings, liked hard paper leaflets, or had concerns about making information posted online look genuine (e.g. if they are posted by official accounts). They often spoke positively about groups dedicated to their ethnic communities (e.g. on Facebook) as illustrated by the following extract:

> "*We now have a new reality. In fact, I suspect that such zoom, zoom meetings will definitely even create a group on a social networking site where you can discuss some posts in a chat, where you post some news, where everyone would have access to it and could read it. You just want to get to what you know, well, through social networks we will reach people the fastest.*" (Marlena, Polish)

Researchers can adjust their activities to make opportunities for people to get involved by providing childcare, and offering reimbursement for peoples' time or arranging meetings at convenient times.

11.5.4.2. Researchers have a trustworthy relationship with the community

This theme emphasises the importance of two-way communication between researchers and the community. Participants argued that in order to create an opportunity for people to become involved, researchers need to establish an ongoing relationship with them. This relationship should expand outside individual research projects and they encouraged researchers to stay in touch (e.g. researchers keeping public contributors up-to-date with their work). Researchers must avoid parachuting themselves into the community, as explained by one of the participants:

> "*An observation from my group administrator, people who create polls, who are looking for people for conversations like you. After getting the data, they leave the group. As if they take something from our group, they find people. They do not leave any feedback, they do not publish anything, they only take people out of the group and my administrator, whom I have known for many years, she talks about how it's a bit unfair that someone comes in, asks for a post about just that they need some people for research, and then leaves no more messages. (…) And this is just such a bit ... In my opinion, a bit unfair. It is not nice to enter the house as a guest, take something and not even say thank you.*" (Mateusz, Polish)

An ongoing relationship with the community can be achieved in many ways, but it takes time. One approach is to work with charities and religious organisations and utilise their leaders as a conduit to reach the community. Involvement and engagement events should not be one-off opportunities to reach the public. However, when appropriately engaged, community, family or religious events can be a great space to build this relationship. It is often researchers who do not know how to involve and engage communities, as presented in the following extract:

> "*When it comes to reaching out and actually reaching our hard to reach communities, I don't think they are really hard to reach. I think there are just organisations that don't know how to reach out to them*." (Greta, Polish)

This theme has shown that trustworthy and ongoing relationship with seldom-heard communities could offer an opportunity for members of the public to get involved. This also means that researchers will understand the community better if they are engaged with them over a longer time period. As explained by the participant:

> “*[It] would be to look into the community and look into their fears and perceptions about that particular study and to address that.”* (Yashica, South Asian)

11.6. Discussion

Our findings show what could influence participants' decision to become public contributors in big data research. However, findings could also be applicable to other health research projects. The COM-B model assisted in mapping what researchers could do to involve Polish and South Asian people/communities in big data research. These experiences were shared by both communities. We identified that there is scope for each factor (capability, opportunity and motivation) to support members of the public to become more engaged in big data research as public contributors. However, not all of these issues could be improved by the researchers themselves. For example, some issues such as being a temporary resident in the UK, is not subject to influence or change by researchers. This can have a significant influence on peoples’ views on involvement and the finding aligns with research around political participation that migrants are less willing to vote if they do not intend to (or know if they will) stay in the country (Salamońska, Lesińska and Kloc-Nowak, 2021). Researchers have a limited scope of influence over peoples’ length of stay (e.g. right to remain) in the UK. This does not mean that they should not mitigate this factor and consider shorter term opportunities or allow public contributors to continue in some capacity even if they leave the UK. Researchers need to make messaging clear to those who consider themselves temporarily in the UK that everyone is welcome no matter how long they intend to stay. It could ensure that public contributors would not feel that their role could not be completed and thus mitigate the challenge of temporality; although, probably not remove it completely. Secondly, researchers should develop trustworthy and ongoing relationship with the community. However, maintaining these relationships might be challenging and time-consuming for researchers (Teodorowski *et al.*, 2023a) Furthermore, researchers can be constrained by funding requirements and might not be able to keep that relationship ongoing. Non-academics might feel that the research process is very slow and even feel frustrated with the timescale (like lengthy publication processes) (Laird *et al.*, 2020).

The COM-B model shows that there can be positive feedback when one enjoys the experience of the behaviour. Our study participants were not existing public contributors. However, many felt they would like to learn more afterwards. They considered the interview experience (and thus research) as positive. Some asked how they could become public contributors; if they did not, the interviewer asked them if they would be interested. In total, fourteen participants received information about becoming public contributors in their local area.

Previous research has shown that there is no single path to achieving successful public involvement (Grotz, Ledgard and Poland, 2020; Brady, 2020). Our study reaffirms that different factors influence the public. However, we would suggest that other researchers consider applying the COM-B model when developing their involvement and engagement strategies. It is possible to map what needs to be put in place to reach particular communities. This would help decide where to shift resources and time for successful involvement, especially as public involvement requires significant investment in this (Boylan *et al.*, 2019).

Our participants supported data reuse under certain conditions. This links to previous research that shows that the public wants to see it only reused for research and policy (Jones *et al.*, 2020a). We found that when researchers follow ethical principles of secure, anonymised and responsible storage, authorised access, and public ownership, our participants felt their data could be reused for research. These principles do not exist in isolation (Waind, 2020). Data security was seen as an issue by some participants in another study exploring the attitudes towards collecting and linking child maltreatment data for research (Moody *et al.*, 2022). Data security emerged as an issue strongly among our participants. However, this could be explained by the lower level of trust in the NHS among some ethnic minorities, which can cause concerns among them about how their data are protected (Bailey Wilson *et al.*, 2022).

We identified multiple factors influencing peoples' willingness to become involved in big data research. Surprisingly, none of our participants directly mentioned the consequences of Brexit (e.g. experience of racism (Rzepnikowska, 2018)) as a challenge for them to be involved. This is despite the fact that there is evidence that Brexit impacted on the integration of European Union citizens in the UK (Teodorowski *et al.*, 2021b).

Some participants felt that they were not experts and thus they should not become involved as public contributors. This challenges the principle of the public offering their lived experience to researchers. However, that finding is not surprising as similar comments were raised by Tend and colleagues (Teng *et al.*, 2019, p. 9), who found that their deliberation participants did not feel that more public involvement was needed around big data research. They themselves "*appreciated the irony*" of that as the process of deliberation was about more public involvement in research. This finding aligns with perspectives of professionals from Mouton Dorey, Baumann and Biller-Andorno (2018) study who felt that patients might not be keen to be involved in governance of big data research and often had limited knowledge of this topic. Some in academia believe that the public can get involved more easily in qualitative research (due to their lived experiences as patients) but not in quantitative because more specialist knowledge is required (Boylan *et al.*, 2019). Even among researchers who support public involvement, not all consider every research stage appropriate for public involvement

(Garfield *et al.*, 2016). Although big data research can be a complex topic, research can (and should) make it inclusive and accessible to members of the public (Aitken *et al.*, 2019).

Education and training are needed for successful public involvement (Manafo *et al.*, 2018), but our study has shown that it might be even more important when involving seldom heard groups in big data research. Members of the public might feel apprehensive about joining big data projects as this is a highly technical area (Jones *et al.*, 2020b). Involving public members with technical skills can be a solution, but the danger is that, as Kirkham and colleagues (Kirkham *et al.*, 2021) point out in their Delphi study, people with big data or research methods experience might have different views than those with less understanding of big data research. We would recommend more promotion of and sessions with the public from seldom-heard groups explaining the role of public contributors and its benefits to the research, researchers and how it can be of benefit to the public contributors themselves.

11.7. Strengths and limitations

Our study is the first to explore how to improve the involvement and engagement of seldom-heard communities in big data research using the COM-B model. However, it has some limitations. It only explored the experiences of Polish and South Asian communities. The COM-B model could guide the exploration of behaviour among other seldom-heard groups, but experiences amongst and within different communities can differ. We would encourage other researchers to explore how the model could be applied to other seldom-heard communities. Also, further research should explore how being a public contributor influences behaviour as our participants were members of the public who have not been previously involved as public contributors. Due to Covid-19 restrictions, we recruited our participants mostly through social media and conducted interviews online; thus, we might have missed the perspectives of people who have less familiarity with technology. Our interview schedule with participants was intentionally broad as there is limited research on how to involve people not yet involved as public contributors around big data research. However, future research could consider using different types of scenarios for discussion. These could include explaining resulting bias that can occur with different forms of consent for big data research. It could also provide examples of what other types of research is conducted, such as on administrative datasets as, on reflection, we felt that many participants limited their discussions to medical research.

11.8. Conclusion

This paper aimed to understand how public involvement in big data research by South Asians and Polish communities in the UK can be encouraged. Using the COM-B model, we identified what measures can be implemented to improve public involvement in these communities. We would encourage researchers to identify what can influence members of the public to be involved and map it under the three factors of capability, opportunity and motivation to determine what measures they could put in place to help reaching and involving seldom-heard communities.

12. Discussion

12.1. Introduction

In this chapter, I bring together findings from three studies and how they relate to the overall aim of this thesis. The aim was to explore how to involve and engage the public (especially seldom-heard communities) in big data research. First, I provide an overview of the research findings and contextualise them in the broader literature. Second, I discuss how this thesis adds to existing knowledge. This includes indicating implications and recommendations arising from this thesis for involving and engaging seldom-heard communities in big data research. Third, I reflect on the strengths and limitations. This is followed by suggestions for future research directions. Then, I reflect on the future of public involvement and engagement in big data research. Last, I share concluding remarks.

12.2. Synopsis of findings

In this thesis, I explored how to involve and engage members of the public (especially from seldom-heard communities) in big data research. My initial premise was confirmed using a scoping review (chapter 5), identifying that there was a knowledge gap, that resulted in limited understanding of how to involve and engage members of the public in big data research, and especially seldom-heard communities. This thesis offered an opportunity to address this gap in the literature. The thesis addressed three research questions, which I discuss in turn:

- RQ 1: How can social media be used to facilitate the engagement process?
- RQ 2: What are the experiences of researchers when involving and engaging seldom-heard communities in big data research?
- RQ 3: How can we involve seldom-heard communities that previously have not been involved in big data research?

To address these research questions, I conducted three studies. The first one looked at engagement on social media through the campaign Data Saves Lives (RQ1). The second study explored the involvement and engagement experiences of researchers (RQ2). The last study examined the perspectives of Polish and South Asian communities (RQ3). The studies comprising this thesis offered novel findings. I provide a synopsis of key findings around each research question and then reflect on how they address the overall aim of the thesis.

12.2.1. RQ1: How can social media be used to facilitate the engagement process?

Twitter data published between April and July 2021 was downloaded using the hashtag #DataSavesLives to explore how the hashtag has been used on Twitter (chapter 9). This international campaign aimed to address public awareness and improve support for big data research (Data Saves Lives, 2020). This was the second study that explored the Data Saves

Lives campaign (Hassan, Nenadic and Tully, 2021), but my study was the first one to explore this topic in a new context because the UK government adopted the motto for its own campaign promoting the GPDPR[59]. Findings showed how the hashtag was used in that time period on Twitter. Overall, these findings confirmed research by Hassan, Nenadic and Tully (2021) that Data Saves Lives was used on Twitter among social media users who were mostly supportive towards the campaign and big data research. The idea of nationwide campaigns to raise awareness of big data research is not new (Goytia *et al.*, 2018; Evans *et al.*, 2020; Sleigh and Vayena, 2021), but the findings suggest that so far, the Data Saves Lives campaign did not fully meet its aims.

The study offered a unique theoretical lens through which to understand how tweets were used within three models of public engagement: deficit, dialogue and participatory. These models were used to identify the relationship and the role of researchers and members of the public in the engagement process. The deficit model is a form of one-way engagement where researchers share information but do not listen to public feedback. The dialogue model is a two-way engagement where both sides communicate with each other. The participatory model also uses two-way engagement but encourages members of the public and researchers to work together to find solutions, thus suggesting co-production. Results from the thematic analysis were mapped under these models. The findings showed that most themes were mapped within the deficit (n=2) and dialogue models (n=2). One theme was mapped within the participatory model. The results of my study did not suggest that one model was better than the other, but it rather recognised these were used simultaneously. This aligns with previous literature suggesting that these models are mixed when used in public engagement (Brossard and Lewenstein, 2010). Metcalfe (2019) showed that these models are not often separated and there is a substantial overlap between them. Still, I would argue that researchers need to understand the different purposes of the models to design a successful public engagement campaign on social media. Therefore, I explore how different engagement messages were used under each model of public engagement.

The deficit model was mostly about communicating that big data research could offer benefits to members of the public and tweets were mostly used to offer a chance to learn more about big data. This indicated that tweets under this model were one-directional. These findings align with previous research by Su *et al.* (2017), showing that science organisations tend to communicate only one way and rarely engage the public. This questions how (and if) tweets under this model actually help the campaign to achieve its aim of raising support for big data

---

[59] For more background about GPDPR, please see at page 53.

research. However, work by Côté, Darling and Heard (2018) suggests that even if members of the public do not engage in any conversation, the message might still reach them. Even if the reach existed in the Data Saves Lives campaign, the previous experience in big data initiatives in the UK indicates that it might not have been enough to gain public support. For example, the care.data[60] project did not gain public support despite reaching over 25 million homes in England through leaflets (Hays and Daker-White, 2015).

The tweets under the dialogue model showed how Twitter users debated sharing data for research and the overarching message was that this process should be ethical. These findings present some insights into the public perspective towards big data research and mostly reaffirmed what is already known in the literature, for example, that the public has concerns about big data research (Kalkman *et al.*, 2019b) or that the public should be engaged to establish trust. This trust could later influence the public to support these new initiatives. This idea of trust aligns with what Carter, Laurie and Dixon-Woods (2015) call social licence.[61] However, what was novel about my research were the suggestions on how to establish this trust. First, a reassurance from researchers that data will not be misused, and that public engagement was seen as one way to do it. This reaffirms Ballantyne and Stewart (2019) argument that engagement could show who could be a trustworthy steward of data. However, it was not clear how to provide this reassurance to the public. Second and as a consequence of trust, there were calls for more public engagement. Twitter users saw it as more two-directional (in contrast to the deficit model), with researchers and the public communicating with each other. That idea of engagement was taken further, as illustrated by the participatory model.

Under the participatory model, more messages appear calling for members of the public to be actively involved in big data research. This could include more public contributors joining big data projects. These findings show that there has been a shift in the discussion on social media about reusing medical data for research. In a previous study, Hays and Daker-White (2015) explored the care.data discussion on Twitter and noticed the lack of discussion on public involvement and engagement. My findings suggest that the discussion on social media has shifted in recent years. This aligns with the work by Bucchi (2008), who argues that campaigns that do not communicate well with members of the public can improve interaction over time.

---

[60] For more background about care.data, please see at page 52.

[61] For more discussion about social licence for big data research, please see at page 55.

12.2.2. RQ2: What are the experiences of researchers when involving and engaging seldom-heard communities in big data research?

Twelve interviews with researchers were conducted to explore their experiences of involving and engaging seldom-heard communities in big data research (chapter 10). This was the first qualitative study that aimed to capture researchers' experiences in this field. Overall, the findings indicate that public involvement and engagement in big data research could be a useful process for both researchers and public contributors. However, findings also could be used to make the case that involvement and engagement in big data research include field-specific obstacles. The findings were presented in four themes that captured participants' experiences. These were: 'abstraction and complexity of big data', 'one size does not fit all', 'working in partnership' and 'empowering the public contribution'. I now discuss how these findings build on previous literature.

The theme 'abstraction and complexity of big data' illustrated the key challenge when involving and engaging members of the public in big data research. Big data consists of technical and complex terms, so participants questioned how this could be discussed in lay terms. This was not a new barrier, as the challenge of the complexity of language and topic has been recognised in the literature on public involvement in big data research (Baart and Abma, 2011; Evans *et al.*, 2020; Manrique de Lara and Peláez-Ballestas, 2020; Deverka *et al.*, 2019; Jones, McNerney and Ford, 2014; Goytia *et al.*, 2018; Dankar, Ptitsyn and Dankar, 2018). However, my findings offered new insights and examples of how to overcome this challenge. Linked with this theme is the 'empowering the public contribution' theme, suggesting that public contributors need to be empowered to be fully able and confident to be involved in big data research. Both themes offered new recommendations to researchers, and I discuss them in detail in another section.[62]

The findings can be used to make an argument that there is no single approach to reaching all communities. Participants spoke about different ways they involved and engaged members of the public. This aligns with previous literature because other researchers recognised that one size does not fit all. For example, when discussing the involvement of people across the lifespan, Wren and CLAPA Cleft Collective Patient Consultation Group (2022, p. 64) recognised that a "*'one size fits all*' approach will not work." Similarly, Harding and Burr (2022, p. 84), reflected on working with children and parents, and recommended that "*one size does not fit all!*". These two works were not involvement activities in big data research specifically

---

[62] See at page 227.

but with seldom-heard communities. This suggests that the findings from the ‘one size does not fit all’ theme might be applicable outside big data research.

The findings in this thesis can be used to recommend that involvement and engagement require working with partner organisations to reach seldom-heard communities. However, participants recognised that identifying potential partners and keeping that relationship can be challenging as it requires time. This contributes to previous research that showed the importance of working with external organisations (e.g. charities) to establish relationships with members of the public (Morrow, Boaz and Sally, 2011; Nelson and Burns, 2020; Steel, 2005a). The findings also offer novel insights into how these partnerships could influence who researchers involve and engage, as one participant pointed out that they inherited diversity from the charity. This question is whether working with external organisations is enough to ensure the broader public is involved and engaged.

My findings offer further insights into which communities are seen as seldom-heard in big data research. However, this seems to depend on the context of the research project, as research participants named different seldom-heard communities in their work. This supports the recommendation made by Paprica *et al.* (2020), who argue that researchers should design an engagement strategy specifically for the community group that might be particularly affected by the planned research. However, my research further indicates that designing that engagement strategy should be done together with the community, as one of my participants described it as “*pre-engagement engagement*”. This strategy could include incorporating already existing customs and traditions (Henare *et al.*, 2019) when working with seldom-heard communities.

The findings from my thesis suggest that big data researchers should be able to support public involvement and engagement. This could lead to practical challenges. First, public involvement and engagement require additional resources, and researchers rarely receive extra time, funds, and training to conduct these activities. This adds extra pressure on them (Boylan *et al.*, 2019), but Garfield *et al.* (2016) do not recognise it as a negative aspect, only one which requires planning. Second, occasionally it can be emotionally burdensome as a patient might share personal, emotionally heavy stories with researchers (Boylan et al., 2019). Therefore, I would argue that funders should expect that adequate PPIE management time is included to support the involvement of the public.

12.2.3. RQ3: How to involve seldom-heard communities that previously have not been involved in big data research?

Forty interviews were conducted with Polish (n=20) and South Asian (n=20) participants to explore how to increase their involvement in big data research (chapter 11). The study had three sub-aims. First, it explored participants' views on how big data are collected and used. Second, it examined enablers and barriers when involving these communities in big data research. Third, the study explored what researchers could do to involve these communities more in big data research. Overall, the findings suggest what might influence the decision by members of the public to become involved (or not) in big data research. However, some findings could also be applicable to public involvement in other research projects. Findings from the thematic analysis were mapped under the COM-B model to identify how different factors could influence participants' decisions to become public contributors. The findings show that different ways exist to involve Polish versus South Asian communities in big data research. This suggests that one size does not fit all communities. Based on this research, I recommend that other researchers use the COM-B model when reflecting on what could influence people in their research. I discuss the application of the COM-B model for public involvement later in this chapter.[63]

One of the key findings was around researchers having a trustworthy (and ongoing) relationship with the community, illustrating how established relationships between the community (e.g. organisations) and researchers could improve the involvement of these communities in big data research. This finding offers a more nuanced understanding of what researchers should do (and should avoid) if they aim for a mutually beneficial relationship. Participants spoke about the problem of researchers parachuting themselves into the community, collecting data or recruiting participants but not giving anything back. This approach risks one-way communication with the community and possible alienation from future research. The finding supports the work by Tindana *et al.* (2015) who argue that establishing a relationship takes time.

The findings contribute to understanding the public's views towards big data research. There is literature on public views around reusing their data, but limited research has been done to capture views of seldom-heard communities. This is the first study exploring Polish community's views. The findings back up previous studies that the public mostly supports big data research but has concerns over confidentiality and data misuse (Kalkman *et al.*, 2019b). For South Asian participants, the findings here are contrary to previous research. Previous

---

[63] See at page 226.

work showed that ethnic minorities could have more concerns than the white British population. For example, Stockdale, Cassell and Ford (2019) suggested that non-white British are less often proponents of data sharing than white British because of their more serious concerns around privacy. Similarly, Bailey Wilson *et al.* (2022) showed that Black and South Asian participants in their study had low trust levels in NHS, and thus, they were more likely to question the level of protection and how their data might be reused. However, here I showed that South Asian (alongside Polish) participants were mostly supportive of data reuse for research. This discrepancy with previous findings indicates the need for further research with these communities.

Another important finding was about the participants' perception of private companies reusing patients' data for research. Previous research mostly showed that the public was concerned about private companies' access to patients' data because there was a perceived potential for data misuse (Hill *et al.*, 2013; Kalkman *et al.*, 2019b; Manhas *et al.*, 2015). The exception was the study by Tully *et al.* (2020). They found that the public would be more supportive of big data research if any profits from big data research were reinvested into public services. My findings build on Tully *et al.* (2020) research and suggest that the public might be inclined to support private companies’ involvement in other instances. Based on findings from my study, I propose that the circumstances when the participants felt comfortable supporting private companies access to patients' medical data would be when the public body (such as a university) is a leading organisation in that research (but has a private company as a partner because of their expertise). Therefore, the findings also challenge Hill *et al.* (2013) study, which looked at acceptable and unacceptable types of research. They found that universities were not always considered the trusted option because their participants questioned who funded the academic work. Interestingly, who pays for research did not come across during my interviews. My findings indicate specific situations when involving private partners in big data research might be acceptable to the public. However, further research should explore if these perspectives differ between size and type of company.

The findings suggest that not all participants perceived themselves as ‘experts by experience’. Some questioned if a member of the public with no expert background should be involved as a public contributor in big data research. They wondered if their knowledge of the topic would be enough to contribute useful advice. This finding reaffirms Mouton Dorey, Baumann and Biller-Andorno (2018) study with professionals, where some believed that the public had limited knowledge and might not be interested in being involved in big data research. It is worth noting that in their study, the participant who was a patient organisation representative had a different opinion. My study findings show that some members of the public might not see their involvement as public contributors in big data research as appropriate or warranted. This raises questions about how the public perceives the role of public contributors.

The majority of participants were unfamiliar with big data terminology. Consequently, this lack of knowledge could influence their ability and willingness to become public contributors. This finding reaffirmed other researchers' position that the technical language (e.g. jargon) could influence if members of the public get involved in big data research (Jones *et al.*, 2020b; Ballantyne and Stewart, 2019). The finding aligns with the researchers' study, where participants also felt that big data is a complex topic that could influence how to involve the public. Thus, I captured this barrier from the perspective of both those who involve and are involved. I discuss some recommendations from both of my studies later in this chapter, providing suggestions for overcoming this challenge.

As participants were from ethnic minorities, the language barrier issue appeared strongly across interviews. One reason this was so prevalent could be that some interviews were conducted in the participants' mother tongue or with the interpreter's support. Thus, participation was easily available for those with limited knowledge of English. Still, the findings suggest that any participant information and public involvement session should be provided in the language the public contributor understands. Recruitment materials should include information on what can be provided for those from seldom-heard communities who might need additional support to participate, e.g. interpreting, childcare, and disability access (Grotz, Ledgard and Poland, 2020). These findings reaffirm the existing language barrier when working with ethnic minorities and migrant communities. For example, Karliner *et al.* (2007), in the systematic review exploring if professional interpreters improve care for patients with limited knowledge of English, found using interpreters seem to raise the quality of care. The participants in my study also suggested the following ways to overcome the language barrier: researchers should speak the participants' language, translate research materials, and have interpreters available.

12.2.4. How do these findings contribute to the research aims?

So far, I have discussed each study's findings and how they answer three distinctive research questions. In this section, I bring the findings together by identifying shared conclusions and show how they answer my overall research question, 'how to involve and engage members of the public (especially from seldom-heard communities) in big data research'.

Together, the three studies make an argument that involvement and engagement in big data research can be carried out in a way that benefits big data research. The findings support the assumptions in the literature that members of the public can be involved in the governance of big data research (Hurt *et al.*, 2019; Ohno-Machado *et al.*, 2014; Milne, Sorbie and Dixon-Woods, 2021; Willison *et al.*, 2019), contribute to big data projects (de Freitas *et al.*, 2021; Funnell *et al.*, 2020), or act as co-researchers (Newburn *et al.*, 2020). However, involvement and engagement should be adequately funded and embedded in big data research (in specific projects or data governance).

Public involvement and engagement in big data research can face challenges. First, all the studies presented here recognised that one approach does not fit all; thus, there can be no one way to reach all communities. The Twitter study demonstrated the diversity of options in how different engagement models were used on Twitter. I suggested that researchers should understand them and appropriately apply them in their public engagement strategies. The researchers and Polish and South Asian communities' studies showed that members of the public could be reached in various ways. These findings put additional pressure on researchers who want to involve and engage seldom-heard communities. They would need to identify new resources and time to design appropriate strategies to reach specific communities. Both the researchers, and the Polish and South Asian participant studies suggest that working in partnership with organisations and charities could facilitate that process. Charities could provide advice and become gatekeepers. Links with the third sector and organisations representing those who researchers try to engage with are crucial in building trust (Steel, 2005b; Morrow, Boaz and Sally, 2011). Little is known about organisations representing and working with seldom-heard communities. These could become gatekeepers and have the potential to act as a partner with public bodies such as NHS in the delivery of services and organising participation strategies or activities (Brady *et al.*, 2020). However, overreliance on gatekeepers could lead to involving people already included with them and potentially missing some sections of the community.

Second, all studies made the case that a challenge is that big data is an abstract and complex topic. Findings from the Twitter study suggest that a social media campaign can be one approach to sharing more information about this field to improve public understanding of the topic. Further practical recommendations arising from researchers' and Polish and South Asian communities' studies are discussed in the next section, focusing on how the thesis contributes to the wider literature on involvement and engagement in big data research.

12.3. How do these findings add to the literature on public involvement and engagement in big data research?

The findings presented in this PhD address knowledge gaps in the existing literature on what is known about the involvement and engagement of seldom-heard communities in big data research. In this section, I discuss this thesis's contributions to the literature. First, I present the new system logic model of public involvement and engagement in big data research that was developed. Second, I explain how the COM-B model for public involvement could be adopted to aid PPI. Last, I offer recommendations based on my findings for people working in big data research and who want to involve and engage members of the public in their work.

12.3.1. System logic model of public involvement and engagement in big data research

The role of system logic models is to guide researchers in considering broad issues that might be of relevance to their research and assist in understanding the complexity of the process. They are particularly popular in reviews (Rohwer *et al.*, 2017). In this section, I discuss the logic model of public involvement and engagement in big data research that I developed in the literature section (chapters 4 and 5). First, I recap the development of the model and reflect on how the studies conducted as part of this thesis further developed it. Then, I discuss how the model can assist researchers who want to unpick the complexity of involving and engaging members of the public in big data research.

The development of the logic model was an iterative process spread over three stages:

1. Following established guidance (Rohwer *et al.*, 2017), the model was developed based on the initial scoping of the literature on public involvement and engagement, discussions with the supervisory team, another doctoral researcher (who was a co-author of the published research protocol) and two public contributors. The initial model is available in chapter 4.[64]
2. Using results from the scoping review, I identified additional aspects of the involvement and engagement process and included them in the logical model. The revised model is available in chapter 5.[65]
3. Further development took place after conducting the three studies in this thesis. In the intervention delivery, I included the facilitators of involvement and engagement process. This provides a balance with the previous point around barriers and encourages researchers to reflect on what they could do to facilitate the involvement and engagement process for members of the public. Figure 11 presents the final logic model.

---

[64] See at page 63.

[65] See at page 90.

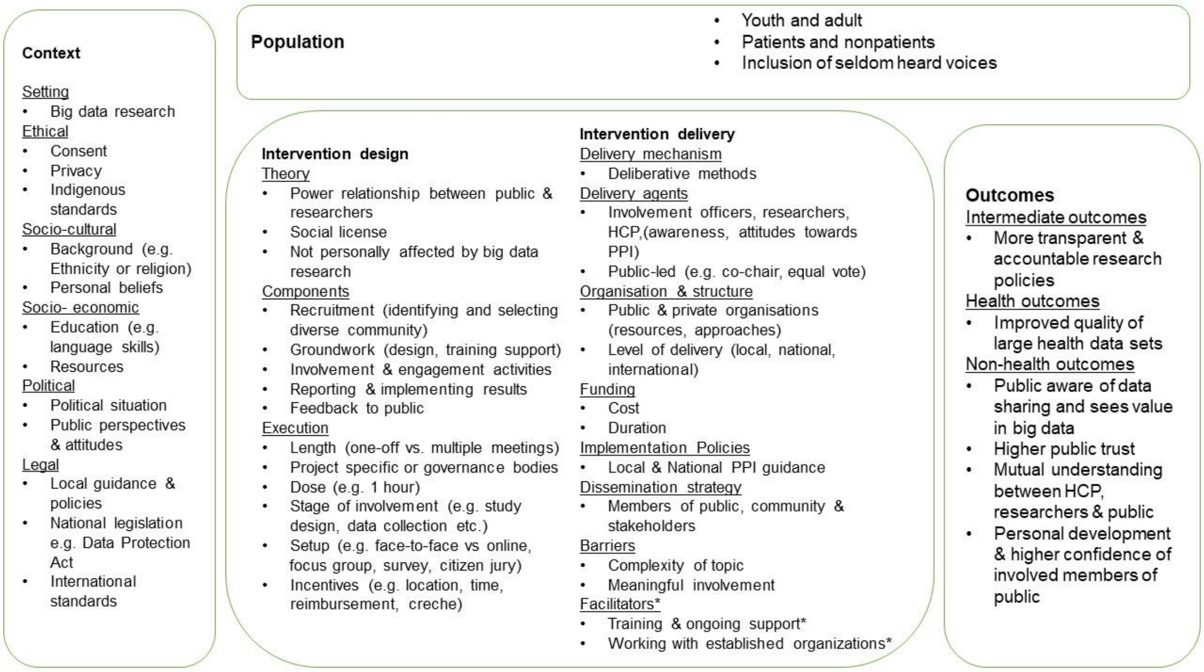


*Figure 11 The final system logic model of public involvement and engagement in big data research. * new aspects of the model based on this thesis.*

The logic model consists of four sections: context, population, intervention (design and delivery) and outcome. I now discuss each of them. The population is about who is a target group to reach through involvement and engagement activities. This recognises the need to reflect on age, reaching both patients and non-patients and inclusion of seldom-heard communities. The context offers space to reflect on what could influence the involvement and engagement activities. These include issues that are related to big data research, such as ethics and legal aspects and what could influence the public perspective, for example, their socio-cultural, socio-economic background and political situation. The intervention provides space to think about all aspects of design and delivery of involvement and engagement activities. It consists of numerous details of what the activities could look like. This is not a prescriptive list but rather a starting point of reflection to recognise how many different paths involvement and engagement activities in big data research could take. Last, the outcomes offer possible consequences of these activities. These can be intermediate, health and non-health related. This is not a full list of possible outcomes but instead aims to assist researchers in identifying what involvement and engagement are supposed to change. This could ensure that the process is not tokenistic but rather benefits researchers and the public.

The logic model is a new way for researchers to plan the design and delivery of involvement and engagement in big data research. The limitation of the model is that it simplifies the process of involvement and engagement. However, it could be used as a baseline for planning these activities and be used alongside theories that offer more insight into particular communities, such as the COM-B model that I explore next.

12.3.2. Adapting the COM-B model to public involvement

Models of behaviour assist researchers in understanding what drives behaviour and how people make decisions. In the Polish and South Asian communities study (chapter 11), I adapted the COM-B model (Michie, Van Stralen and West, 2011) as it offered a coherent framework to understand behaviour. In my case, it was about exploring what influences participants' decisions to become (or not) public contributors in big data research. Consequently, I saw an individual decision to become involved as a public contributor to be a behaviour that could be influenced. In this section, I discuss further the COM-B model and how it could apply to public involvement projects.

The COM-B model suggests that three interacting factors are required for behaviour to take place (or to be changed). These are capability, opportunity and motivation. Capability makes a behaviour possible to occur. It has two components: psychological and physical. In the case of public involvement, it is more the former and depends on whether people have the knowledge, skills and abilities required to be involved as public contributors in big data research. This includes people’s perspectives on whether they have the necessary skills and knowledge. Some participants in Polish and South Asian communities’ study questioned whether they had enough understanding of big data research to be involved. Therefore, researchers need to provide that capacity and assist people in realising that they have it. This could be achieved by raising awareness about what role of public contributor involves and the importance of lived experience in research. The perspective that one does not have capability might influence the next component of the model: motivation.

Motivation is about mental processes that influence and lead behaviour reflectively or automatically. Public involvement depends on whether there is motivation to influence a person’s decision to become a public contributor. For example, it can consist of conscious thought processes like plans and evaluations and beliefs about their capabilities and consequences or more instinctive processes like desires and habits, optimism, and emotions. In terms of big data research, the involvement should become something that a public contributor is interested in doing to become more active. Motivation might be around the willingness to learn new skills around research, but this could be restricted by limited finances, which are linked to the next factor, opportunity.

Opportunity makes it possible for the behaviour to take place. In public involvement, these can be both physical (e.g. right resources, accessible location and time) and social opportunities (culture and social norms) to make public involvement in research possible. For example, limited financial resources are one of the barriers for people to become involved (Snow, Tweedie and Pederson, 2018). Even if opportunities for members of the public to become public contributors in big data research exist, people need to be aware of them and encouraged to be involved in them. Limited resources can cause public contributors opportunities to be mostly fulfilled by people with more time and resources, e.g. retirees and the middle class.

These factors are interlinked and influence each other and the resulting behaviour, creating a feedback loop within the model. Capability and opportunity influence motivation. Higher capability and opportunity for involvement lead to improved motivation and, consequently, a better chance of behaviour change. For example, people might become involved in research because they feel motivated to contribute their lived experiences as patients or carers. They wish to share these experiences, often to improve healthcare services (Staniszewska *et al.*, 2007; Lauzon-Schnittka *et al.*, 2022; Paul *et al.*, 2020). This could be a challenge regarding big data research as members of the public might not feel it directly relates to their lived experiences. Conversely, if someone does not feel motivated, they might not want to be involved (even if invited to become a public contributor). Behaviour also can influence all factors in both positive and negative ways. In theory, when having a good experience as a public contributor, people should feel more willing to continue in their role or sign up for a similar opportunity. Conversely, if one had a bad experience, that might drive them not to do it again (with the same or another research team).

The Polish and South Asian communities' paper was the first study adopting the COM-B model in public involvement research. This offered a novel way to understand what could influence members of the public to become public contributors. Researchers could consider adopting this approach when developing their own strategies to involve new public contributors, especially those coming from seldom-heard communities. This approach can be applicable outside big data research, but more research is needed to evaluate it.

12.3.3. Recommendations for involving and engaging seldom-heard communities in big data research

The findings from this thesis offer a new understanding of public involvement and engagement, specifically in seldom-heard communities. In this section, I present key recommendations for researchers and other professionals working in big data which involve and engage the public. This build and expands on recommendations already made in previous chapters that report the research findings. First, I discuss recommendations on involvement and engagement in big data research. Second, I reflect on some recommendations for working with seldom-heard communities as these might be applicable outside big data and be of relevance to other types of health research.

Recommendations for involving and engaging members of the public in big data research:

1. Providing training and information about big data research to ensure everyone who wants to be involved has enough understanding of the topic to contribute in a way that benefits the research. Current public contributors should receive ongoing support from researchers and research institutions.
2. Researchers who want to involve and engage members of the public in their work need buy-in from their research organisation. This includes additional time, resources for these activities and training on how to do public involvement and engagement. Therefore, public involvement and engagement need to be embedded in the research culture. This embedment should be an ongoing process, not a one-off opportunity.
3. PPIE can be resource-intensive. Some big data research projects and institutions could consider a PPIE coordinator/manager to facilitate public involvement. This would release researcher time and support them with appropriate PPIE expertise.
4. Due to ongoing training and developing research skills in big data, a public contributor might start to offer a more expert perspective than a lay view. Therefore, their role in governance bodies should be time-limited.
5. Public engagement should be two-way and transparent communication, with researchers giving feedback to members of the public on how they reflected and responded to their comments. However, researchers should receive guidance on how to do it. This could be achieved by developing appropriate policies at the institutional level.

The first recommendation builds on the premise that members of the public might have various and often limited knowledge of big data research, so they should be offered more information about it. This does not equal education but rather provides opportunities to learn more if they want. Educating the public could be seen as a paternalistic approach (Shaw, Sethi and Cassel, 2020; Willison *et al.*, 2019), where researchers explain but do not learn themselves from the public what they think or why this is the case. Moreover, existing public contributors can require ongoing support. This could include explaining technical terms in lay language or detailed discussion about complex research processes to allow public contributors to make them able and comfortable to be involved.

The second recommendation is about embedding public involvement and engagement in a research environment so every researcher would have an opportunity to include public contributors in their work. This builds on De Simoni *et al.* (2023) paper arguing that the research environment needs more people with experience in public involvement in research. They further claim that funding applications often underestimate the cost of public involvement, especially how much staff time and resources it takes. Funders should expect costs for researchers to manage PPIE to be added to the applications. Still, I agree with Ballantyne and Stewart (2019) argument that engagement activities should be proportional to the type and size of big data initiatives.

The third recommendation is directly related to the second one and argues for creating a PPIE coordinator/manager role to support and facilitate public involvement and engagement activities. This new role aligns with previous research arguing that people who conduct PPIE should be trained in doing so (Hurt *et al.*, 2019; Teng *et al.*, 2019). Training on how to involve and engage the public should be available both to the researchers and members of the public (Jones *et al.*, 2020b). However, the creation of PPIE-specific roles should not exempt big data researchers from actively involving and engaging members of the public in their work but rather offer them support on how to do it.

The fourth recommendation builds on the first one. It recognises that if a public contributor stays a long time on a research project or governance board, they might start to offer an expert view. Public contributors on big data projects often offer lay perspective rather than the lived experience. This does not mean that their contribution is not valid. However, it might be questioned if a public contributor still represents a lay perspective after being involved in big data research and receiving training on methodology. One might question if they offer public or professional perspective. Based on the researchers' study, I suggested that three years would be an appropriate time for a public contributor to hold an advisory role, and this aligns with previous research. For example, O'Doherty *et al.* (2011) propose the term for the governance bodies for public contributors to be between two to three years.

The fifth recommendation calls for two-way communication. What researchers share with the public should be relevant, of interest and understandable by both sides. The public has a right to expect their concerns and comments to be taken on board. Researchers should provide feedback explaining how they used these comments. However, if it is not possible to implement their recommendations in practice, it should be explained why.

Recommendations for involving and engaging seldom-heard communities in big data that might be applicable to other health research.

1. For each project, researchers should identify specific seldom-heard communities relevant to their work. The focus could be on members of the public who share multiple characteristics to provide further and more diverse insights.
2. Establish an ongoing and trustworthy relationship with the community before and after the project ends. This means that researchers need to avoid parachuting themselves into the community.
3. Work with organisations that already include seldom-heard communities but also keep involvement and engagement opportunities open to people not involved with these established groups.
4. PPI is a novel concept for some and many may not know what a public contributor is, especially if there is no equivalent in their mother tongue. Information in accessible language should be available on the role of the public contributor.
5. Provide opportunities for people on the move (e.g. migrant communities). This could include temporary public involvement roles and chances to remain involved after moving away (e.g. by remote involvement).

The first recommendation is based on the researchers' study, where participants discussed various seldom-heard communities in the context of their involvement and engagement in big data research. Therefore, at the start of a project, researchers should identify which communities would be most relevant for their work. This recommendation reaffirms NIHR (2020) guidance that seldom-heard communities remain "highly context-specific".

The second recommendation suggests having an ongoing relationship between the community and researchers could support researchers in reaching seldom-heard communities. This statement would not be new. However, I further argue that this relationship needs to be mutually beneficial to both sides. For example, researchers should avoid appearing only when they need a piece of advice from public contributors or new data. Again, this puts additional pressure and commitments on researchers, so they should be provided with more support (such as a PPIE manager) and resources to do it.

The third recommendation links directly with the previous one as it urges researchers to work with organisations but recognises the danger of limiting themselves to people who these organisations already include.

The fourth recommendation recognises that public involvement is a new concept for some members of the public. There is a need to actively promote and raise awareness of public involvement if researchers want to involve people not previously included. This could consist of meetings, workshops or leaflets raising awareness of public involvement in health research.

The last recommendation adds to the fourth one as it is about offering flexibility for public contributors who consider themselves living temporarily in the area (or the UK, like migrant communities) to become involved. New public contributors' opportunities should include a clear statement that this flexibility is welcomed.

This thesis offered a unique understanding of how to involve and engage seldom-heard communities in big data research from the perspective of researchers, and Polish and South Asian communities.
Building on previous social media research, the Twitter study demonstrated the potential of social media to engage members of the public in big data research initiatives. It identified how social media messages align with public engagement models (deficit, dialogue and participatory).
For the first time, the COM-B model was suggested as a way for researchers to reflect on what could influence members of the public decision to become public contributors.
The scoping review proposed a logic model of public involvement and engagement in big data research. This is a unique contribution because all aspects of the involvement and engagement process have been mapped.
This doctoral research suggested new approaches to involve and engage seldom-heard communities in big data research.

Table 12 The summary of the thesis's contributions to the field.

This section focused on how this thesis contributes to the broader literature on public involvement and engagement in big data research. First, I proposed two models that can support researchers that want to involve and engage members of the public in big data research. Then, I made recommendations arising from this research. These were about the involvement and engagement of members of the public in big data research and how to involve and engage seldom-heard communities. Based on this discussion and the previous section, where I offered a synopsis of findings, I summarise this thesis' novel contributions to the field in table 12. Despite this thesis offering new findings in the field, my research also has limitations. Therefore, the next section focuses on the strengths and limitations of this thesis.

12.4. Strengths and limitations of this doctoral research

In this section, I reflect on the strengths and limitations of this thesis. This discussion builds on already recognised strengths and limitations in previous chapters that reported the research findings. However, I do not recap the strengths and limitations of each individual study but rather reflect on overlapping themes and issues. First, I reflect on the overall strength of this thesis. Second, I reflect on the challenge of how participants might understand public involvement, engagement and big data research. Then, I reflect on the generalisability of my findings. Last, I briefly discuss the alternative direction this doctoral research could have taken.

Public involvement and engagement in big data research remains a developing field. This thesis offered novel findings that contribute to the literature in this area. These included the first empirical evidence on this topic. The researchers' study (chapter 10) was the first project that looked at this from the perspective of researchers who had experience in involving and engaging seldom-heard communities in big data research. The Polish and South Asian communities (chapter 11) were the first explorations of how to involve these two communities and, according to my knowledge, also the first study to specifically look at migrant and ethnic minority groups in the UK in big data research. The Twitter study (chapter 9) was a follow-up work on how the social media campaign “Data Saves Lives” has evolved. Exploring these different perspectives of researchers, members of the public and social media discourse was a strength of this thesis as it provided diverse views into public involvement and engagement in big data research. However, as participation was voluntary, it is possible that only researchers who were supportive of public involvement and engagement took part. Researchers with less experience in public involvement and engagement or who hold negative views toward it were likely to have had different views.

There are many seldom-heard communities, and it is not possible to provide an exhaustive list. Polish and South Asian participants are a sample of seldom-heard communities, and further research should explore other seldom-heard groups, as other groups might have different views.

As much as these studies offered a glimpse into how to involve and engage seldom-heard communities in big data research, the findings could not be used to determine the impact of these activities. First, it could not be confirmed how successful (if at all) these activities could be in building trust between researchers and communities. Second, it remained unclear if public involvement and engagement actually make the public more supportive of big data research.

The timing of the Twitter study was one of its strengths, as data was collected during the public controversy in England surrounding the new big data initiative (GPDPR). Interviews with researchers, Polish and South Asian participants took place within a year after the completion of the Twitter study. Interestingly, participants did not speak much about the public controversies around big data research and did not directly mention the GPDPR. On reflection, I could have asked participants directly about the impact of these events. For researchers, this would be if and how the controversy influenced their involvement and engagement plans or activities. For Polish and South Asian participants, if they heard about this and what were their thoughts about it.

In the researchers' study, I provided a short definition of public involvement and engagement to ensure the participants and I had the same understanding during the interviews. This was particularly relevant with participants based outside the UK; for example, one participant was from Canada, and they mostly use the term public engagement for what in the UK is called involvement. A similar strategy was used for Polish and South Asian community studies. I included a short description of big data research for participants. However, many participants asked for more examples during the discussion and after the interview, so the guide included further examples with the link to the Data Saves Lives website (Data Saves Lives, 2023). This was written in lay language. The decision to provide only a short overview was guided by time restrictions as the study consisted of one-hour interviews conducted online. This was a good basis for discussion with participants. However, some participants thought about health research as research on medical research rather than broader public health. One explanation behind this could be that traditionally members of the public become involved because of their own or family health conditions rather than public health issues (e.g. benefits of green space to mental health).

Another limitation of my research is how much findings are generalisable to other contexts, situations and populations. It is important to emphasise that my thesis aimed to present multiple perspectives on involving and engaging the public in big data research. However, capturing the views of all seldom-heard communities and other relevant parties in big data research was impossible (and indeed was not intentionally pursued as qualitative research does not attempt to be generalisable). These could include public contributors already involved in big data researchers, big data professionals and funders. Second, participant recruitment was undertaken during the Covid-19 pandemic, which could have influenced participation from certain backgrounds, for example, people with good digital skills. Therefore, findings cannot be generalisable to all seldom-heard communities. Nevertheless, as some findings overlapped between participants within and between studies, they might be applicable when working with other seldom-heard communities.

This doctoral research could have taken other directions. In the researchers' study, I recognised that one of the limitations was excluding the perspective of the public contributors. While this research provided some valuable insights into the involvement and engagement process, I could have considered using a case study approach and included, alongside researchers' interviews, members of the public who were involved and engaged in the projects that participants spoke about. This would have offered a comparison of views and experiences. However, due to the Covid-19 pandemic, it was challenging to establish relationships and identify research institutions that would be open to allow that level of access to both groups. Another direction that I considered was interviewing public contributors who were already involved in big data research. However, after an initial discussion with big data researchers, I recognised that seldom-heard communities are rarely included in their involvement and engagement activities. Thus, I decided to focus on the experience of members of the public not previously included as public contributors.

The section provided a critical reflection on the strength and limitations of this research. The latter would suggest some further research which I discuss in the next section.

12.5. Future directions for public involvement and engagement in big data research

This section explores avenues for further research in the involvement and engagement of seldom-heard communities in big data research. However, I do not revisit the recommendations already made in published papers but rather offer overarching directions for future research in the field.

This thesis explored how to involve Polish and South Asian communities and the experiences of researchers in involving seldom-heard communities. However, what constitutes seldom-heard communities is broad; further research should focus on other communities to see if similar themes emerge from other seldom-heard communities, or whether there are additional specific considerations that might be of relevance to that specific community. Some members of the public might have more concerns than the general public about big data research. For example, people might feel that their personal circumstances would make them more identifiable in datasets, such as LGBT+ communities, or people living with HIV. Further qualitative work would allow for capturing these perspectives. These could use different methods. I recognised in the previous section that one of the limitations was the interview length. Future research should consider utilising longer sessions (e.g. half-day focus groups). This could offer more nuanced insights into the topic and allow participants to exchange viewpoints. It could take an approach similar to Hill *et al.* (2013), who conducted focus groups during which they explained why researchers prefer opt-out rather than opt-in. Then, they explored participants' views based on that presentation. This provided a new understanding of how public perspective could change when well-informed about the research process.

One study in this thesis looked at the perspectives of researchers and the other on members of the public. Future research could bring these two groups together. Focus groups could be an appropriate method for both groups to exchange experiences and reflect on disagreements. Moreover, if the data collected for this thesis is used alongside these focus groups, this could offer further insights into the involvement and engagement process. For example, findings from the researchers’ study could be presented to public contributors and used as a base for discussion.

The researchers' study offered further examples of how to involve and engage seldom-heard communities in big data research. This added to the current literature. Future research should focus on developing new materials on public involvement and engagement that big data researchers could utilise when working with members of the public. Moreover, examples of involvement and engagement activities in big data research have rarely been evaluated. Future research into and evaluation of the process of involvement and engagement is needed. One way to do it is to use the four-dimensional framework to map and evaluate the experience of public contributors (Gibson, Welsman and Britten, 2017). This could also consist of longitudinal studies that would bring independent evaluators alongside the research team or use co-production approaches. The NIHR (2023) recognised this gap by opening a funding stream focusing on evaluating the public acceptance of administrative data for research. This includes an evaluation of the effectiveness of public engagement activities in shaping public views. The need for evaluation also applies to the COM-B model of behaviour that I suggested in the Polish and South Asian communities’ study as a way for researchers to reflect on how to involve and engage seldom-heard communities in big data research. This would confirm if the COM-B model is applicable to other public involvement projects.

In the Twitter study paper, I suggested that future research should explore how the Data Saves Lives campaign keeps evolving on Twitter and how it is being used on Facebook and Instagram. However, Data Saves Lives is not the only way to promote and engage members of the public around big data research on social media. Researchers and organisations often use social media to engage the public. These individual case studies should be explored. The particular focus could be on Twitter accounts run by certain research teams or organisations focusing on big data research (for example, Administrative Data Research UK that tweets as @adr_uk).

12.6. Future of public involvement and engagement in big data research

Public involvement and engagement have become embedded in health research. This thesis provided further insights into how to involve and engage members of the public (especially from seldom-heard communities) in big data research. However, at the same time, findings indicate some challenges in making it happen. First, big data can be a complex and abstract topic for members of the public. Second, there is no singular way to involve and engage all communities. This thesis suggested two new ways of thinking about involvement and engagement in big data research. First, the system logic model for public involvement and engagement in big data research is a comprehensive way for researchers to think about the whole process of involving and engaging members of the public. Second, I suggested that researchers who wish to involve seldom-heard communities use the COM-B model to map what could facilitate people's willingness to become public contributors. These models accompany already existing standards and guidance in big data research and broader health research. I argue that these should be used alongside my suggested models to fully benefit researchers who want to involve and engage members of the public (especially seldom-heard communities) in big data research. In this section, I explain which standards and guidance align well with the models applied in this thesis. First, I look at existing ones, and then I discuss current developments in the field. Last, I reflect if embedding involvement and engagement in big data research is feasible in the future.

Two guidance documents about involvement and engagement are currently available to researchers working in big data field. First, outside big data research, researchers could refer to the National Standards for Public Involvement (NIHR, 2019). These provide guidance on what 'good' involvement should look like. Second, the consensus statement on public involvement and engagement with data-intensive research (Aitken *et al.*, 2019) offers field-specific guidance. The statement names eight principles for involvement and engagement in big data research. Table 13 presents these standards and principles. The guidance is not mandatory but offers best practices for working with the members of the public.

National Standards for Public Involvement (NIHR, 2019)	Key principles for public involvement and engagement in data-intensive health research as adopted from Aitken et al. (2019)
1. Inclusive opportunities	1. Have institutional buy-in
2. Working together	2. Have clarity of purpose
3. Support & learning	3. Be transparent
4. Communications	4. Involve two-way communication
5. Impact	5. Be inclusive and accessible to broad publics
6. Governance	6. Be ongoing
	7. Be designed to produce impact
	8. Be evaluated

Table 13 Principles and standards for public involvement and engagement

The thesis confirms these principles' relevance for big data research as most of them came across in the findings. Furthermore, the recommendations arising from this thesis expanded on the guidance. The Consensus statement called for more support for members of the public and for diversity

> "*Those involved in [involvement and engagement] activities should be supported to make the most of the engagement opportunities offered and to freely and fully articulate their views. [Involvement and engagement] should facilitate the participation of diverse groups and interests*" (Aitken *et al.*, 2019, p. 5).

My research has built on these principles and provides insights and examples for researchers on achieving this in practice, especially coming from seldom-heard communities. The National Standards for Public Involvement specify inclusive opportunities that there is a need to "*identify and address barriers to taking up public involvement in research*" (NIHR, 2019, p. 6). Although there is no comprehensive list of what these challenges might be for each community, the standards provide some suggestions. This thesis offered a new way of identifying and reflecting on them by suggesting the use of the COM-B model when developing involvement opportunities. Big data researchers could use these established principles alongside my research findings when considering how recommendation arising from this thesis could influence their work.

Big data research is still developing, so further work on involving and engaging the public has been taking place. A new organisation, Research Data Scotland, was established in 2022 to facilitate involvement and engagement with the public in Scotland in big data research. Research Data Scotland joined the newest alliance, Public Engagement in Data Research Initiative (PEDRI, 2023), a sector-wide partnership that enables organisations working with big data to bring collective actions to embed involvement and engagement in their work and others'. This is a UK wide-initiative launched in 2022 that, for the first time, brings together research organisations, funders and regulators. One of their main aims is to create public involvement and engagement best practice standards for data-driven research and statistics. I have been involved in developing these standards as a member of the working group and shared some findings from this thesis. There is a clear overlap between some recommendations from this thesis and the standards. First, PEDRI recognises that big data research is a complex topic and suggests that members of the public should have a baseline knowledge of it to be involved as public contributors. Second, it argues for the two-way communication between researchers and members of the public that continues after the research comes to an end. Third, PEDRI argues for embedding involvement and engagement culture in organisations working with big data research. Fourth, there is a call for involving and engaging all communities. The standards are seen as a starting point of the discussion and could be a helpful resource for researchers working in big data research. This resource could potentially be used alongside the models I suggest in this thesis to ensure involvement and engagement that benefits both researchers and members of the public. At the time of writing this thesis, public consultations for these standards have taken place, and the finalised standards should be published in due course.

These recent developments show growing institutional support for involvement and engagement and embedding these activities in big data research. This thesis's findings have contributed to this positive change. PEDRI standards, alongside this thesis's findings and recommendations, have the potential to support researchers and members of the public further. This indicates that involvement and engagement in big data research might become more prominent in the future. This would mean that researchers would become more comfortable organising and conducting these activities. Moreover, members' of the public knowledge and understanding of that field would improve and thus also their ability to be engaged as well as involved as public contributors.

12.7. Concluding remarks

This thesis offered a new body of evidence on how to involve and engage members of the public, especially coming those from seldom-heard communities, in big data research. Findings contributed to existing public involvement and engagement literature by drawing on discussions on how to involve and engage members of the public in big data research. This was achieved by exploring the perspectives of researchers, Polish and South Asian communities and how discussion occurs on social media. Findings suggest that there is no one approach to reaching a variety of seldom-heard communities, and researchers need to adapt and develop appropriate strategies for each community. This could be achieved by working with partner organisations. Findings imply that barriers exist to public involvement and engagement with seldom-heard communities in big data research. These are especially about the use of complex, technical language. The recommendations from this thesis could benefit researchers who want to involve and engage seldom-heard communities in big data research and could encourage funders to ensure that appropriate funding is provided to make it happen. The application of the system logic model of public involvement and engagement in big data research and the COM-B model will be particularly applicable to those who want to better understand how to improve their involvement and engagement strategies. Last, this thesis proposed avenues for further research in the field.

Word count: 91613

References

Abbasi, K. (2022) 'The red flag patient data breaches of a wannabe global data superpower', *BMJ*, pp. o1174.

Adamek, M. E. (2015) 'Building Scholarly Writers: Student Perspectives on Peer Review in a Doctoral Writing Seminar', *Journal of Teaching in Social Work,* 35(1-2), pp. 213-225.

Aitken, M., Cunningham-Burley, S. and Pagliari, C. (2016) 'Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme', *Science and Public Policy,* 43(5), pp. 713-723.

Aitken, M., de St. Jorre, J., Pagliari, C., Jepson, R. and Cunningham-Burley, S. (2016) 'Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies', *BMC Medical Ethics,* 17(1), pp. 73.

Aitken, M., Porteous, C., Creamer, E. and Cunningham-Burley, S. (2018) 'Who benefits and how? Public expectations of public benefits from data-intensive health research', *Big Data & Society,* 5(2), pp. 2053951718816724.

Aitken, M., Tully, M., Porteous, C., Denegri, S., Cunningham-Burley, S., Banner, N., Black, C., Burgess, M., Cross, L., Delden, J., Ford, E., Fox, S., Fitzpatrick, N., Gallacher, K., Goddard, C., Hassan, L., Jamieson, R., Jones, K., Kaarakainen, M. and Willison, D. (2019) 'Consensus Statement on Public Involvement and Engagement with Data-Intensive Health Research', *International Journal of Population Data Science,* 4.

Alexander, K. E., Brijnath, B. and Mazza, D. (2014) 'Barriers and enablers to delivery of the Healthy Kids Check: an analysis informed by the Theoretical Domains Framework and COM-B model', *Implementation Science,* 9(1), pp. 60.

Andits, P. (2016) 'Between Suspicion and Trust: Fieldwork in the Australian-Hungarian Community', *Insider Research on Migration and Mobility*: Routledge, pp. 135-152.

Archibald, M. M., Ambagtsheer, R. C., Casey, M. G. and Lawless, M. (2019) 'Using Zoom Videoconferencing for Qualitative Data Collection: Perceptions and Experiences of Researchers and Participants', *International Journal of Qualitative Methods,* 18, pp. 160940691987459.

Arksey, H. and O'Malley, L. (2005) 'Scoping studies: towards a methodological framework', *International Journal of Social Research Methodology,* 8(1), pp. 19-32.

Armstrong, S. (2021) 'GP data scheme may face further delay as practices refuse to hand over patient information', *BMJ,* 374, pp. n1777.

Arnstein, S. R. (1969) 'A Ladder Of Citizen Participation', *Journal of the American Institute of Planners,* 35(4), pp. 216-224.

Baart, I. L. M. A. and Abma, T. A. (2011) 'Patient participation in fundamental psychiatric genomics research: a Dutch case study', *Health Expectations,* 14(3), pp. 240-249.

Bailenson, J. N. (2021) 'Nonverbal overload: A theoretical argument for the causes of Zoom fatigue.', *Technology, Mind, and Behavior,* 2(1).

Bailey, J. (2008) 'First steps in qualitative data analysis: transcribing', *Family Practice,* 25(2), pp. 127-131.

Bailey Wilson, B., Twins, B., Wilkinson-Salamea, C., Raidos, D., Imafidon, K., McGarry, N., Sheni, R. and Hicks, S. (2022) *A participatory research project: Exploring the views and experiences of Black and South Asian communities in the UK on patient data and its uses*: ClearView Research. Available at: https://understandingpatientdata.org.uk/sites/default/files/2022-04/Diverse%20voices%20on%20Data%20-%20Main%20report_0.pdf (Accessed: 19 October 2022).

Ballantyne, A. and Stewart, C. (2019) 'Big Data and Public-Private Partnerships in Healthcare and Research', *Asian Bioethics Review,* 11(3), pp. 315-326.

Ballantyne, A. and Style, R. (2017) 'Health data research in New Zealand: Updating the ethical governance framework', *New Zealand Medical Journal,* 130(1464), pp. 64-71.

Baltar, F. and Brunet, I. (2012) 'Social research 2.0: virtual snowball sampling method using Facebook', *Internet Research,* 22(1), pp. 57-74.

Barker, F., Atkins, L. and De Lusignan, S. (2016) 'Applying the COM-B behaviour model and behaviour change wheel to develop an intervention to improve hearing-aid use in adult auditory rehabilitation', *International Journal of Audiology,* 55(sup3), pp. S90-S98.

Becker, S., Sempik, J. and Bryman, A. (2010) 'Advocates, Agnostics and Adversaries: Researchers' Perceptions of Service User Involvement in Social Policy Research', *Social Policy and Society,* 9(3), pp. 355-366.

Beier, K., Schweda, M. and Schicktanz, S. (2019) 'Taking patient involvement seriously: a critical ethical analysis of participatory approaches in data-intensive medical research', *BMC Medical Informatics and Decision Making,* 19(1).

Beresford, P. (2002) 'User Involvement in Research and Evaluation: Liberation or Regulation?', *Social Policy and Society,* 1(2), pp. 95-105.

Beresford, P. (2003) 'User involvement in research: Exploring the challenges', *NT Research,* 8(1), pp. 36-46.

Beresford, P. (2012) 'The Theory and Philosophy Behind User Involvement', in Beresford, P. and Carr, S. (eds.) *Social care, service users and user involvement*. London: Jessica Kingsley Publishers.

Beresford, P. (2013) *Beyond the usual suspects: Towards inclusive user involvement: Findings.* Shaping Our Lives Publ.

Beresford, P. (2019a) 'Including Our Self In Struggle: Challenging the neo-liberal psycho-system's subversion of us, our ideas and action', *Canadian Journal of Disability Studies,* 8(4), pp. 31-59.

Beresford, P. (2019b) 'Public Participation in Health and Social Care: Exploring the Co-production of Knowledge', *Frontiers in Sociology,* 3.

Beresford, P. (2020) 'PPI Or User Involvement: Taking stock from a service user perspective in the twenty first century', *Research Involvement and Engagement,* 6(1).

Beresford, P. (2021) *Participatory ideology: From exclusion to involvement.* Policy Press.

Berger, P., L and Luckmann, T. (1966) 'The social construction of reality', *A Treatise in the Sociology of Knowledge*.

Berry, N., Lobban, F., Belousov, M., Emsley, R., Nenadic, G. and Bucci, S. (2017) '#WhyWeTweetMH: Understanding Why People Use Twitter to Discuss Mental Health Problems', *Journal of Medical Internet Research,* 19(4), pp. e107.

Beyer, K. M., Comstock, S. and Seagren, R. (2010) 'Disease maps as context for community mapping: a methodological approach for linking confidential health information with local geographical knowledge for community health research', *J Community Health,* 35(6), pp. 635-44.

Bharti, N., O'Donovan, C., Smallman, M. and Wilson, J. (2021) 'Public Trust, Deliberative Engagement and Health Data Projects: Beyond Legal Provisions', *Engaging Science, Technology, and Society,* 7(1), pp. 125-133.

Bloom, L. R. and Sawin, P. (2009) 'Ethical responsibility in feminist research: challenging ourselves to do activist research with women in poverty', *International Journal of Qualitative Studies in Education,* 22(3), pp. 333-351.

BMA and RCGP (2021) *General Practice Data for Planning and Research - communications with the public*. Available at: https://www.rcgp.org.uk/-/media/Files/News/rcgp-bma-letter-gpdpr.ashx?la=en (Accessed: 10 January 2022).

BMJ Open (2021) *Peer Review History*. Available at: https://bmjopen.bmj.com/content/bmjopen/11/8/e050167.reviewer-comments.pdf (Accessed: 20 February 2023).

Boden, C., Edmonds, A. M., Porter, T., Bath, B., Dunn, K., Gerrard, A., Goodridge, D. and Stobart, C. (2021) 'Patient partners' perspectives of meaningful engagement in synthesis reviews: A patient-oriented rapid review', *Health Expectations*.

Bogen, K. W., Mulla, M. M. M., Haikalis, M. and Orchowski, L. M. (2022) 'Sexual Victimization Among Men: A Qualitative Analysis of the Twitter Hashtag #UsToo', *Journal of Interpersonal Violence,* 37(9-10), pp. NP7825-NP7849.

Boghossian, P. A. (2001) 'What is social construction?', *Times Literary Supplement,* (February), pp. 6-8.
Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., Brozek, I. and Hughes, C. (2014) 'Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups', *BMC Medical Research Methodology,* 14(1), pp. 42.
Boote, J., Wong, R. and Booth, A. (2015) ''Talking the talk or walking the walk?' A bibliometric review of the literature on public involvement in health research published between 1995 and 2009', *Health Expectations,* 18(1), pp. 44-57.
Bot, B. M., Wilbanks, J. T. and Mangravite, L. M. (2019) 'Assessing the consequences of decentralizing biomedical research', *Big Data and Society,* 6(1).
Boyatzis, R. E. (1998) *Transforming qualitative information: Thematic analysis and code development.* sage.
Boylan, A. M., Locock, L., Thomson, R. and Staniszewska, S. (2019) '"About sixty per cent I want to do it": Health researchers' attitudes to, and experiences of, patient and public involvement (PPI)—A qualitative interview study', *Health Expectations,* 22(4), pp. 721-730.
Boyle, D., Coote, A., Sherwood, C. and Slay, J. (2010) *Right Here, Right Now. Taking co-production into the mainstream*. Available at: https://media.nesta.org.uk/documents/right_here_right_now.pdf (Accessed: 15 December 2020).
Bradby, H., Green, G., Davison, C. and Krause, K. (2017) 'Is Superdiversity a Useful Concept in European Medical Sociology?', *Frontiers in Sociology,* 1.
Bradshaw, A. S. (2022) '#DoctorsSpeakUp: Exploration of Hashtag Hijacking by Anti-Vaccine Advocates and the Influence of Scientific Counterpublics on Twitter', *Health Communication*, pp. 1-11.
Brady, L.-M. (2020) 'Rhetoric to reality: the need for a new approach', in Brady, L.-M. (ed.) *Embedding Young People's Participation in Health Services. New Approaches*. Bristol: Policy Press, pp. 249-266.
Brady, L.-M., Roberts, E., Hathway, F. and Horn, L. (2020) 'Shifting sands: trying to embed participation in a climate of change', in Brady, L.-M. (ed.) *Embedding Young People's Participation in Health Services. New Approaches*. Bristol: Policy Press, pp. 177-201.
Braun, V. (2008) '"She'll be right"? National identity explanations for poor sexual health statistics in Aotearoa/New Zealand', *Social Science & Medicine,* 67(11), pp. 1817-1825.
Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology,* 3(2), pp. 77-101.
Braun, V. and Clarke, V. (2016) '(Mis)conceptualising themes, thematic analysis, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis', *International Journal of Social Research Methodology,* 19(6), pp. 739-743.
Braun, V. and Clarke, V. (2019) 'Reflecting on reflexive thematic analysis', *Qualitative Research in Sport, Exercise and Health,* 11(4), pp. 589-597.
Braun, V. and Clarke, V. (2021a) 'Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches', *Counselling and Psychotherapy Research,* 21(1), pp. 37-47.
Braun, V. and Clarke, V. (2021b) 'One size fits all? What counts as quality practice in (reflexive) thematic analysis?', *Qualitative Research in Psychology,* 18(3), pp. 328-352.
Braun, V. and Clarke, V. (2021c) *Thematic Analysis: A Practical Guide.* London: Sage.
Braun, V. and Clarke, V. (2021d) 'To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales', *Qualitative Research in Sport, Exercise and Health,* 13(2), pp. 201-216.
Braun, V. and Clarke, V. (2023) 'Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting', *Health Psychology Review*, pp. 1-24.

Braun, V., Clarke, V. and Hayfield, N. (2022) ''A starting point for your journey, not a map': Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis', *Qualitative Research in Psychology,* 19(2), pp. 424-445.

Braun, V., Clarke, V., Hayfield, N. and Terry, G. (2019) 'Thematic Analysis': Springer Singapore, pp. 843-860.

Brett, J., Staniszewska, S., Mockford, C., Herron-Marx, S., Hughes, J., Tysall, C. and Suleman, R. (2014) 'Mapping the impact of patient and public involvement on health and social care research: a systematic review', *Health expectations : an international journal of public participation in health care and health policy,* 17(5), pp. 637-650.

Brickman Bhutta, C. (2012) 'Not by the Book:Facebook as a Sampling Frame', *Sociological Methods & Research,* 41(1), pp. 57-88.

Brooks, J., McCluskey, S., Turley, E. and King, N. (2015) 'The Utility of Template Analysis in Qualitative Psychology Research', *Qual Res Psychol,* 12(2), pp. 202-222.

Broomfield, K. (2022) 'The Unspoken Voices Project: Creating accessible research through collaborative design', in Volkmer, A. and Broomfield, K. (eds.) *Seldom Heard Voices in Service User Involvement*. Croydon: J&R Press, pp. 111-128.

Broomfield, K., Craig, C., Smith, S., Jones, G., Judge, S. and Sage, K. (2021) 'Creativity in public involvement: supporting authentic collaboration and inclusive research with seldom heard voices', *Research Involvement and Engagement,* 7(1).

Brossard, D. and Lewenstein, B. (2010) 'A Critical Appraisal of Models of Public Understanding of Science: Using Practice to Inform Theory', in Kahlor, L. and Stout, P. (eds.) *Communicating Science: New Agendas in Communication*. New York: Routledge, pp. 11-39.

Bryman, A. (2016) *Social research methods.* Oxford university press.

Bucchi, M. (2008) 'Of deficits, deviations and dialogues: Theories of public communication of science', *Handbook of public communication of science and technology*: Routledge, pp. 71-90.

Burns, T. W., O'Connor, D. J. and Stocklmayer, S. M. (2003) 'Science Communication: A Contemporary Definition', *Public Understanding of Science,* 12(2), pp. 183-202.

Burr, V. (2015) *Social Constructivism.* London: Routledge.

Byrne, D. (2022) 'A worked example of Braun and Clarke's approach to reflexive thematic analysis', *Quality & Quantity,* 56(3), pp. 1391-1412.

Carling, J., Erdal, M. B. and Ezzati, R. (2014) 'Beyond the insider–outsider divide in migration research', *Migration Studies,* 2(1), pp. 36-54.

Carney, L., Jones, L., Braddon, F., Pullyblank, A. M. and Dixon, A. R. (2006) 'A Colorectal Cancer Patient Focus Group Develops an Information Package', *The Annals of The Royal College of Surgeons of England,* 88(5), pp. 447-449.

Carter, P., Laurie, G. T. and Dixon-Woods, M. (2015) 'The social licence for research: why care.data ran into trouble', *Journal of Medical Ethics,* 41(5), pp. 404.

Carter, P. and Martin, G. (2016) 'Challenges Facing Healthwatch, a New Consumer Champion in England', *International Journal of Health Policy and Management,* 5(4), pp. 259-263.

Charmaz, K. (2000) 'Grounded theory: Objectivist and constructivist methods', *Handbook of qualitative research,* 2(1), pp. 509-535.

Charmaz, K. (2006) *Constructing grounded theory: A practical guide through qualitative analysis.* sage.

Charmaz, K. and Belgrave, L. (2012) 'Qualitative interviewing and grounded theory analysis', *The SAGE handbook of interview research: The complexity of the craft,* 2, pp. 347-365.

Clark, M., Vliet, E. and Collins, M. (2021) 'Reflections from the COVID-19 pandemic on inequalities and patient and public involvement and engagement (PPIE) in social care, health and public health research', *Health Expectations,* 24(5), pp. 1547-1550.

Clarke, V. and Braun, V. (2018) 'Using thematic analysis in counselling and psychotherapy research: A critical reflection', *Counselling and Psychotherapy Research,* 18(2), pp. 107-110.

Clarke, V., Braun, V. and Hayfield, N. (2015) 'Thematic analysis', *Qualitative psychology: A practical guide to research methods,* 222(2015), pp. 248.
Cochrane (2017) *Consumer involvement in Cochrane - the Statement of Principles*. Available at: https://consumers.cochrane.org/sites/consumers.cochrane.org/files/public/uploads/statement_of_principles_for_consumer_involvement_in_cochrane_july_2017.pdf (Accessed: 12 January 2021).
Cochrane (2020) *A network for patients and carers within Cochrane*. Available at: https://consumers.cochrane.org/healthcare-users-cochrane (Accessed: 5 December 202).
Colquhoun, H. L., Levac, D., O'Brien, K. K., Straus, S., Tricco, A. C., Perrier, L., Kastner, M. and Moher, D. (2014) 'Scoping reviews: time for clarity in definition, methods, and reporting', *Journal of Clinical Epidemiology,* 67(12), pp. 1291-1294.
Conrad, P. and Barker, K. K. (2010) 'The social construction of illness: Key insights and policy implications', *Journal of health and social behavior,* 51(1_suppl), pp. S67-S79.
Corbin, J. and Strauss, A. (1988) *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage publications.
Costa, D. S. J., Mercieca-Bebber, R., Tesson, S., Seidler, Z. and Lopez, A.-L. (2019) 'Patient, client, consumer, survivor or other alternatives? A scoping review of preferred terms for labelling individuals who access healthcare across settings', *BMJ Open,* 9(3), pp. e025166.
Côté, I. M., Darling, E. S. and Heard, S. B. (2018) 'Scientists on Twitter: Preaching to the choir or singing from the rooftops?', *FACETS,* 3(1), pp. 682-694.
Cotterell, P. and Paine, M. (2012) 'Involving a Marginalized Group in Research and Analysis - People with Life Limiting Conditions. Issues and Gains', in Beresford, P. and Carr, S. (eds.) *Social Care, Service Users and User Involvement*. London: Jessica Kingsley Publishers.
Cottrell, E. K., Whitlock, E. P., Kato, E., Uhl, S., Belinson, S., Chang, C., Hoomans, T., Meltzer, D. O., Noorani, H., Robinson, K. A., Motu'apuaka, M., Anderson, J., Paynter, R. A. and Guise, J.-M. (2015) 'Defining the benefits and challenges of stakeholder engagement in systematic reviews', *Comparative Effectiveness Research,* 5, pp. 13-19.
Coulter, A. (2011) *Engaging Patients in Health Care.*
Coulter, A. (2021) 'Patient trust in plans to share primary care data', *BMJ : British Medical Journal (Online),* 373.
Cowley, A., Kerr, M., Darby, J. and Logan, P. (2019) 'Reflections on qualitative data analysis training for PPI partners and its implementation into practice', *Research Involvement and Engagement,* 5(1).
Cox, M. and Ellsworth, D. 'Application-controlled demand paging for out-of-core visualization'. *Proceedings. Visualization'97 (Cat. No. 97CB36155)*: IEEE, 235-244.
Craig, E. (2021) 'Matt Hancock says 'data saves lives' and dismisses idea of sending letters to patients to tell them how they can opt out of controversial NHS data-sharing plan because 'hardly anyone writes to you these days'', *The Daily Mail*. Available at: https://www.dailymail.co.uk/news/article-9712079/Matt-Hancock-dismisses-idea-writing-patients-tell-opt-NHS-data-sharing.html (Accessed: 29 July 2022).
Crawford, M. J. (2002) 'Systematic review of involving patients in the planning and development of health care', *BMJ,* 325(7375), pp. 1263-1263.
Creswell, J. W. (2017) *Research design : Qualitative, quantitative, and mixed methods approaches / John W. Creswell. Core textbook* 5th edition (international student edition). edn. London: SAGE Publications.
Crouch, H. (2021) *New data collection service could 'destroy' patient trust*. Available at: https://www.digitalhealth.net/2021/05/data-collection-service-destroy-patient-trust/ (Accessed: 15 February 2023).
Dankar, F. K., Ptitsyn, A. and Dankar, S. K. (2018) 'The development of large-scale de-identified biomedical databases in the age of genomics-principles and challenges', *Human Genomics,* 12(1).
Data Saves Lives (2020) *What is Data Saves Lives?* Available at: https://datasaveslives.eu/aboutdsl (Accessed: 11 January 2021).

Data Saves Lives (2023) *Case Summaries*. Available at: https://datasaveslives.eu/casesummaries (Accessed: 20 June 2023).
Dawson, S., Campbell, S. M., Giles, S. J., Morris, R. L. and Cheraghi-Sohi, S. (2018) 'Black and minority ethnic group involvement in health and social care research: A systematic review', *Health Expectations,* 21(1), pp. 3-22.
Dawson, S., Ruddock, A., Parmar, V., Morris, R., Cheraghi-Sohi, S., Giles, S. and Campbell, S. (2020) 'Patient and public involvement in doctoral research: reflections and experiences of the PPI contributors and researcher', *Research Involvement and Engagement,* 6(1).
de Freitas, C., Amorim, M., Machado, H., Elisa Leão, T., Maria João, B., Renedo, A., Provoost, V. and Silva, S. (2021) 'Public and patient involvement in health data governance (DATAGov): protocol of a people-centred, mixed-methods study on data use and sharing for rare diseases care and research', *BMJ Open,* 11(3).
De Simoni, A., Jackson, T., Inglis Humphrey, W., Preston, J., Mah, H., Wood, H. E., Kinley, E., Gonzalez Rienda, L. and Porteous, C. (2023) 'Patient and public involvement in research: the need for budgeting PPI staff costs in funding applications', *Research Involvement and Engagement,* 9(1).
Deakin, H. and Wakefield, K. (2014) 'Skype interviewing: reflections of two PhD researchers', *Qualitative Research,* 14(5), pp. 603-616.
Department of Health (1989) *Working for Patients. White Paper*.
Department of Health (1991) *Patient's Charter*.
Department of Health (2005) *Creating a Patient-led NHS. Delivering the NHS Improvement Plan*. Available at: https://www.plymouth.ac.uk/uploads/production/document/path/1/1921/dh_4106507.pdf (Accessed: 25 January 2021).
Department of Health (2006) *Best Research for Best Health*.
Department of Health (2012) *Local Healthwatch: A strong voice for people - the policy explained*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215097/dh_133288.pdf (Accessed: 20 January 2021).
Department of Health & Social Care (2022a) *Data saves lives: reshaping health and social care with data*. Available at: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data (Accessed: 2 August 2022).
Department of Health & Social Care (2022b) *Data saves lives: reshaping health and social care with data (draft)*. Available at: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data-draft/data-saves-lives-reshaping-health-and-social-care-with-data-draft (Accessed: 1 August 2022).
Department of Health and Social Care (2012) *NHS Constitution for England*.
Deverka, P. A., Gilmore, D., Richmond, J., Smith, Z., Mangrum, R., Koenig, B. A., Cook-Deegan, R., Villanueva, A. G., Majumder, M. A. and McGuire, A. L. (2019) 'Hopeful and Concerned: Public Input on Building a Trustworthy Medical Information Commons', *J Law Med Ethics,* 47(1), pp. 70-87.
Dixon-Woods, M. and Ashcroft, R. E. (2008) 'Regulation and the social licence for medical research', *Medicine, Health Care and Philosophy,* 11(4), pp. 381-391.
Dixon, S., Agha, K., Ali, F., El-Hindi, L., Kelly, B., Locock, L., Otoo-Oyortey, N., Penny, S., Plugge, E. and Hinton, L. (2018) 'Female genital mutilation in the UK- where are we, where do we go next? Involving communities in setting the research agenda', *Research Involvement and Engagement,* 4(1).
Domecq, J. P., Prutsky, G., Elraiyah, T., Wang, Z., Nabhan, M., Shippee, N., Brito, J. P., Boehmer, K., Hasan, R., Firwana, B., Erwin, P., Eton, D., Sloan, J., Montori, V., Asi, N., Abu Dabrh, A. M. and Murad, M. H. (2014) 'Patient engagement in research: a systematic review', *BMC Health Services Research,* 14(1), pp. 89.

Duchange, N., Darquy, S., d'Audiffret, D., Callies, I., Lapointe, A. S., Loeve, B., Boespflug-Tanguy, O. and Moutel, G. (2014) 'Ethical management in the constitution of a European database for leukodystrophies rare diseases', *Eur J Paediatr Neurol,* 18(5), pp. 597-603.

Eccles, A., Bryce, C., Turk, A. and Atherton, H. (2018) 'Patient and public involvement mobile workshops – convenient involvement for the un-usual suspects', *Research Involvement and Engagement,* 4(1).

Edley, N. (2001) 'Unravelling Social Constructionism', *Theory & Psychology,* 11(3), pp. 433-441.

Edwards, D. (1997) *Discourse and cognition.* Sage.

Edwards, R. (1998) 'A critical examination of the use of interpreters in the qualitative research process', *Journal of Ethnic and Migration Studies,* 24(1), pp. 197-208.

Erikainen, S., Friesen, P., Rand, L., Jongsma, K., Dunn, M., Sorbie, A., McCoy, M., Bell, J., Burgess, M., Chen, H., Chico, V., Cunningham-Burley, S., Darbyshire, J., Dawson, R., Evans, A., Fahy, N., Finlay, T., Frith, L., Goldenberg, A., Hinton, L., Hoppe, N., Hughes, N., Koenig, B., Lignou, S., McGowan, M., Parker, M., Prainsack, B., Shabani, M., Staunton, C., Thompson, R., Varnai, K., Vayena, E., Williams, O., Williamson, M., Chan, S. and Sheehan, M. (2021) 'Public involvement in the governance of population-level biomedical research: unresolved questions and future directions', *Journal of Medical Ethics,* 47(7), pp. 522-525.

European Patients Forum (2019) *'DATA SAVES LIVES' UNVEILED AT EPF CONGRESS*. Available at: https://www.eu-patient.eu/news/latest-epf-news/2019/data-saves-lives-unveiled-at-epf-congress/ (Accessed: 2 December 2021).

Evans, E. A., Delorme, E., Cyr, K. and Goldstein, D. M. (2020) 'A qualitative study of big data and the opioid epidemic: recommendations for data governance', *BMC Medical Ethics,* 21(1).

Fernando, B., King, M. and Sumathipala, A. (2019) 'Advancing good governance in data sharing and biobanking - international aspects', *Wellcome Open Research,* 4.

Fischbacher, C. M., Hunt, S. and Alexander, L. (2004) 'How physically active are South Asians in the United Kingdom? A literature review', *Journal of Public Health,* 26(3), pp. 250-258.

Fleurence, R. L., Beal, A. C., Sheridan, S. E., Johnson, L. B. and Selby, J. V. (2014) 'Patient-powered research networks aim to improve patient care and health research', *Health Aff (Millwood),* 33(7), pp. 1212-9.

Flynn, R., Walton, S. and Scott, S. D. (2019) 'Engaging children and families in pediatric Health Research: a scoping review', *Research Involvement and Engagement,* 5(1), pp. 32.

Foley, L., Kiely, B., Croke, A., Larkin, J., Smith, S. M., Clyne, B., Pierce, M., Murphy, E. and Team, C. D. A. i. M. (2021) 'A protocol for the evaluation of the process and impact of embedding formal and experiential Public and Patient Involvement training in a structured PhD programme', *Journal of Multimorbidity and Comorbidity,* 11, pp. 26335565211024793.

Foot, C., Gilburt, H., Dunn, P., Jabbal, J., Seale, B., Goodrich, J., Buck, D. and Taylor, J. (2014) 'People in control of their own health and care', *King's Fund*.

Ford, E., Boyd, A., Bowles, J. K. F., Havard, A., Aldridge, R. W., Curcin, V., Greiver, M., Harron, K., Katikireddi, V., Rodgers, S. E. and Sperrin, M. (2019) 'Our data, our society, our health: A vision for inclusive and transparent health data science in the United Kingdom and beyond', *Learning Health Systems,* 3(3), pp. e10191.

Francis, R. (2013) *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*. Available at: https://webarchive.nationalarchives.gov.uk/20150407084231/http://www.midstaffspublicinquiry.com/report.

Freimuth, V. S. and Mettger, W. (1990) 'Is there a hard-to-reach audience?', *Public health reports (Washington, D.C. : 1974),* 105(3), pp. 232-238.

Fuller, M. Y. and Allen, T. C. (2016) 'Let's Have a Tweetup: The Case for Using Twitter Professionally', *Archives of Pathology & Laboratory Medicine,* 140(9), pp. 956-957.

Funnell, S., Tanuseputro, P., Letendre, A., Bearskin, L. B. and Walker, J. (2020) 'Nothing about us, without us. How community-based participatory research methods were adapted in an indigenous end-of-life study using previously collected data', *Canadian Journal on Aging,* 39(2), pp. 145-155.

Galbin, A. (2014) 'An introduction to social constructionism', *Social research reports,* 6(26), pp. 82-92.

Gale, N. K., Heath, G., Cameron, E., Rashid, S. and Redwood, S. (2013) 'Using the framework method for the analysis of qualitative data in multi-disciplinary health research', *BMC Medical Research Methodology,* 13(1), pp. 117.

Gallier, S., Price, G., Pandya, H., McCarmack, G., James, C., Ruane, B., Forty, L., Crosby, B. L., Atkin, C., Evans, R., Dunn, K. W., Marston, E., Crawford, C., Levermore, M., Modhwadia, S., Attwood, J., Perks, S., Doal, R., Gkoutos, G., Dormer, R., Rosser, A., Fanning, H. and Sapey, E. (2021) 'Infrastructure and operating processes of PIONEER, the HDR-UK Data Hub in Acute Care and the workings of the Data Trust Committee: A protocol paper', *BMJ Health and Care Informatics,* 28(1).

Garfield, S., Jheeta, S., Husson, F., Jacklin, A., Bischler, A., Norton, C. and Franklin, B. D. (2016) 'Lay involvement in the analysis of qualitative data in health services research: a descriptive study', *Research Involvement and Engagement,* 2(1).

Garrison, N. A., Sathe, N. A., Antommaria, A. H., Holm, I. A., Sanderson, S. C., Smith, M. E., McPheeters, M. L. and Clayton, E. W. (2016) 'A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States', *Genet Med,* 18(7), pp. 663-71.

Gawlewicz, A. (2016) 'Language and translation strategies in researching migrant experience of difference from the position of migrant researcher', *Qualitative Research,* 16(1), pp. 27-42.

Gergen, K. J. (1973) 'Social psychology as history', *Journal of personality and social psychology,* 26(2), pp. 309.

Gergen, K. J. (1989) 'Warranting voice and the elaboration', *Texts of identity*, pp. 70-81.

Gergen, K. J. (2015) *An invitation to social construction.* 3 edn.: Sage.

Gibson, A., Welsman, J. and Britten, N. (2017) 'Evaluating patient and public involvement in health research: from theoretical model to practical workshop', *Health Expectations,* 20(5), pp. 826-835.

Gilbertson, A., Ridd, M. J., Sutton, E., Liddiard, L., Clayton, J., Roberts, A., Chan, J., Bhanot, A., Wellesley, R. and Dawson, S. (2021) 'Engaging with diverse audiences to raise awareness about childhood eczema: reflections from two community events', *Research Involvement and Engagement,* 7(1).

Glasby, J. and Beresford, P. (2006) 'Commentary and Issues : Who knows best? Evidence-based practice and the service user contribution', *Critical Social Policy,* 26(1), pp. 268-284.

Goldacre, B. and Morley, J. (2022) *Better, Broader, Safer: Using Health Data for Research and Analysis*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1067053/goldacre-review-using-health-data-for-research-and-analysis.pdf (Accessed: 27 July 2022).

Gonzalez, E. and Lincoln, Y. (2006) 'Decolonizing Qualitative Research: Non-traditional Reporting Forms in the Academy', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research,* 7.

Gorman, D. R., Bielecki, K., Willocks, L. J. and Pollock, K. G. (2019) 'A qualitative study of vaccination behaviour amongst female Polish migrants in Edinburgh, Scotland', *Vaccine,* 37(20), pp. 2741-2747.

Gough, D., Thomas, J. and Oliver, S. (2012) 'Clarifying differences between review designs and methods', *Systematic Reviews,* 1(1), pp. 28.

Goytia, C. N., Kastenbaum, I., Shelley, D., Horowitz, C. R. and Kaushal, R. (2018) 'A Tale of 2 Constituencies: Exploring Patient and Clinician Perspectives in the Age of Big Data', *Med Care,* 56 Suppl 10 Suppl 1(10 Suppl 1), pp. S64-s69.

Grant, A., Maciver, E., Adams, N., Teodorowski, P. and Kennedy, C. (2023) '‘A journey of self-discovery and transformation’: A theoretical and comprehensive evaluation of the Queen's nursing institute Scotland community development programme', *Journal of Advanced Nursing,* n/a(n/a).

Gray, L., Wong, G., Rempel, G. and Cook, K. (2020) 'Expanding Qualitative Research Interviewing Strategies: Zoom Video Communications', *Qualitative Report,* 25, pp. Article 9.

Grotz, J., Ledgard, M. and Poland, F. (2020) *Patient and Public Involvement in Health and Social Care Research.* Springer International Publishing.

Gubrium, E. and Koro-Ljungberg, M. (2005) 'Contending With Border Making in the Social Constructionist Interview', *Qualitative Inquiry,* 11(5), pp. 689-715.
Gubrium, J. F. and Holstein, J. A. (2008) 'The constructionist mosaic', *Handbook of constructionist research.–2008.–P*, pp. 3-12.
Guest, G., Bunce, A. and Johnson, L. (2006) 'How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability', *Field Methods,* 18(1), pp. 59-82.
Gunningham, N., Kagan, R. A. and Thornton, D. (2004) 'Social License and Environmental Protection: Why Businesses Go Beyond Compliance', *Law & Social Inquiry,* 29(2), pp. 307-341.
Hanafin, J. and Lynch, A. (2002) 'Peripheral Voices: Parental involvement, social class, and educational disadvantage', *British Journal of Sociology of Education,* 23(1), pp. 35-49.
Harding, S. and Burr, S. (2022) 'Working with children and parents', in Volkmer, A. and Broomfield, K. (eds.) *Seldom Heard Voices in Service User Involvement: The how and why of meaningful collaboration*: J & R Press Limited, pp. 73-86.
Harrison, J. D., Anderson, W. G., Fagan, M., Robinson, E., Schnipper, J., Symczak, G., Carnie, M. B., Hanson, C., Banta, J., Chen, S., Duong, J., Wong, C. and Auerbach, A. D. (2019a) 'Patient and Family Advisory Councils for Research', *JONA: The Journal of Nursing Administration,* 49(10), pp. 473-479.
Harrison, J. D., Auerbach, A. D., Anderson, W., Fagan, M., Carnie, M., Hanson, C., Banta, J., Symczak, G., Robinson, E., Schnipper, J., Wong, C. and Weiss, R. (2019b) 'Patient stakeholder engagement in research: A narrative review to describe foundational principles and best practice activities', *Health Expectations,* 22(3), pp. 307-316.
Harrison, S., Dowswell, G. and Milewa, T. (2002) 'Guest editorial: public and user 'involvement' in the UK National Health Service', *Health and Social Care in the Community,* 10(2), pp. 63-66.
Hassan, L., Nenadic, G. and Tully, M. P. (2021) 'A Social Media Campaign (#datasaveslives) to Promote the Benefits of Using Health Data for Research Purposes: Mixed Methods Analysis', *Journal of Medical Internet Research,* 23(2), pp. e16348.
Hay, A. and Samra-Fredericks, D. (2016) 'Desperately seeking fixedness: Practitioners' accounts of 'becoming doctoral researchers'', *Management Learning,* 47(4), pp. 407-423.
Hays, R. and Daker-White, G. (2015) 'The care.data consensus? A qualitative analysis of opinions expressed on Twitter', *BMC Public Health,* 15(1).
Hazell, C. M., Chapman, L., Valeix, S. F., Roberts, P., Niven, J. E. and Berry, C. (2020) 'Understanding the mental health of doctoral researchers: a mixed methods systematic review with meta-analysis and meta-synthesis', *Systematic Reviews,* 9(1).
HDR UK (2021) *Call to action: #DataSavesLives - We all need to be represented*. Available at: https://www.hdruk.ac.uk/news/data-saves-lives-we-all-need-to-be-represented-call-to-action/#utm_source=hive&utm_medium=email (Accessed: 15 November 2021).
Health and Social Care Committee, House of Commons (2021) *Inquiry: General Practice Data for Planning and Research*.
Hemingway, H., Asselbergs, F. W., Danesh, J., Dobson, R., Maniadakis, N., Maggioni, A., Van Thiel, G. J. M., Cronin, M., Brobert, G., Vardas, P., Anker, S. D., Grobbee, D. E. and Denaxas, S. (2018) 'Big data from electronic health records for early and late translational cardiovascular research: challenges and potential', *European Heart Journal,* 39(16), pp. 1481-1495.
Hemming, L., Pratt, D., Bhatti, P., Shaw, J. and Haddock, G. (2021) 'Involving an individual with lived-experience in a co-analysis of qualitative data', *Health Expectations*.
Henare, K. L., Parker, K. E., Wihongi, H., Blenkiron, C., Jansen, R., Reid, P., Findlay, M. P., Lawrence, B., Hudson, M. and Print, C. G. (2019) 'Mapping a route to Indigenous engagement in cancer genomic research', *Lancet Oncology,* 20(6), pp. e327-e335.
Hennink, M. and Kaiser, B. N. (2022) 'Sample sizes for saturation in qualitative research: A systematic review of empirical tests', *Social Science & Medicine,* 292, pp. 114523.
Hill, E. M., Turner, E. L., Martin, R. M. and Donovan, J. L. (2013) '"Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study', *BMC Medical Research Methodology,* 13(1), pp. 72.

Hodson, H. (2016a) *Did Google's NHS patient data deal need ethical approval?*: New Scientist. Available at: https://www.newscientist.com/article/2088056-did-googles-nhs-patient-data-deal-need-ethical-approval/ (Accessed: 18 February 2023).

Hodson, H. (2016b) *Revealed: Google AI has access to huge haul of NHS patient data*: New Scientist. Available at: https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/ (Accessed: 18 February 2023).

Hogg, C. (2009) *Citizens, Consumers and the NHS. Capturing Voices.* Basingstoke: Palgrave Macmillan.

Hong, Q., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C. and Vedel, I. (2018) *Mixed Methods Appraisal Tool (MMAT)*. Available at: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf (Accessed: 12/11 2020).

Howe, N., Giles, E., Newbury-Birch, D. and Mccoll, E. (2018) 'Systematic review of participants' attitudes towards data sharing: a thematic synthesis', *Journal of Health Services Research & Policy,* 23(2), pp. 123-133.

HRA (2021) *Public involvement in a pandemic: Lessons from the UK COVID-19 public involvement matching service*. Available at: https://s3.eu-west-2.amazonaws.com/www.hra.nhs.uk/media/documents/8948_Public_Involvement_in_Pandemic_Research_Report_V9_-_Accessible.pdf (Accessed: 13 January 2021).

Hsieh, H.-F. and Shannon, S. E. (2005) 'Three Approaches to Qualitative Content Analysis', *Qualitative Health Research,* 15(9), pp. 1277-1288.

Hudson, M., Garrison, N. A., Sterling, R., Caron, N. R., Fox, K., Yracheta, J., Anderson, J., Wilcox, P., Arbour, L., Brown, A., Taualii, M., Kukutai, T., Haring, R., Te Aika, B., Baynam, G. S., Dearden, P. K., Chagné, D., Malhi, R. S., Garba, I., Tiffin, N., Bolnick, D., Stott, M., Rolleston, A. K., Ballantyne, L. L., Lovett, R., David-Chavez, D., Martinez, A., Sporle, A., Walter, M., Reading, J. and Carroll, S. R. (2020) 'Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data', *Nature Reviews Genetics,* 21(6), pp. 377-384.

Hui, A. and Stickley, T. (2007) 'Mental health policy and mental health service user perspectives on involvement: a discourse analysis', *Journal of Advanced Nursing,* 59(4), pp. 416-426.

Hurt, L., Ashfield-Watt, P., Townson, J., Heslop, L., Copeland, L., Atkinson, M. D., Horton, J. and Paranjothy, S. (2019) 'Cohort profile: HealthWise Wales. A research register and population health data platform with linkage to National Health Service data sets in Wales', *BMJ Open,* 9(12).

Iacobucci, G. (2017) 'Patient data were shared with Google on an "inappropriate legal basis," says NHS data guardian', *BMJ,* 357, pp. j2439.

INVOLVE (2012a) *Diversity and inclusion: What's it about and why is it important for public involvement in research?* , Diversity and inclusion: What's it about and why is it important for public involvement in research?

INVOLVE (2012b) *Public involvement in systematic reviews: Supplement to the briefing notes for researchers*, Eastleigh: INVOLVE. Available at: https://www.invo.org.uk/posttypepublication/public-involvement-in-systematic-reviews/.

INVOLVE (2012c) *Strategies for diversity and inclusion in public involvement: Supplement to the briefing notes for researchers.* Available at: https://www.invo.org.uk/wp-content/uploads/2012/06/INVOLVEInclusionSupplement1.pdf (Accessed: 20 November 2020).

INVOLVE (2020) *What is public involvement in research?* Available at: https://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/ (Accessed: 21 September 2020).

IPFCC (2018) *Diverse Voices Matter: Improving Diversity in Patient and Family Advisory Councils*. Available at: https://ipfcc.org/resources/Diverse-Voices-Matter.pdf.

Islam, S. and Small, N. (2020) 'An annotated and critical glossary of the terminology of inclusion in healthcare and health research', *Research Involvement and Engagement,* 6(1), pp. 14.

Ives, J., Damery, S. and Redwod, S. (2013) 'PPI, paradoxes and Plato: who's sailing the ship?', *Journal of Medical Ethics,* 39(3), pp. 181-185.
Jackson, S. J. and Foucault Welles, B. (2015) 'Hijacking #myNYPD: Social Media Dissent and Networked Counterpublics', *Journal of Communication,* 65(6), pp. 932-952.
Jewell, A., Pritchard, M., Barrett, K., Green, P., Markham, S., Mckenzie, S., Oliver, R., Wan, M., Downs, J. and Stewart, R. (2019) 'The Maudsley Biomedical Research Centre (BRC) data linkage service user and carer advisory group: creating and sustaining a successful patient and public involvement group to guide research in a complex area', *Research Involvement and Engagement,* 5(1).
Jones, B. and Hunt, A. (2022) 'Collaboration between doctoral researchers and patient research partners: reflections and considerations', *Research for All,* 6(1), pp. 1-9.
Jones, E., Frith, L., Gabbay, M., Tahir, N., Hossain, M., Goodall, M., Bristow, K. and Hassan, S. (2022a) 'Remote working in public involvement: findings from a mixed methods study', *Research Involvement and Engagement,* 8(1).
Jones, K., Heys, S., Thompson, R., Griffiths, L., Johnson, R., Lee, A., Ford, D. and Broadhurst, K. (2020a) 'Linking pre- and post-adoption records for research in anonymised form in a data safe haven:' *International Journal of Population Data Science,* 5(3).
Jones, K. H., Ford, D. V., Thompson, S. and Lyons, R. A. (2019) 'A profile of the SAIL databank on the UK secure research platform', *International Journal of Population Data Science,* 4(2).
Jones, K. H., Heys, S., Thompson, R., Cross, L. and Ford, D. (2020b) 'Public involvement & engagement in the work of a data safe haven: A case study of the SAIL Databank', *International Journal of Population Data Science,* 5(3).
Jones, K. H., Laurie, G., Stevens, L., Dobbs, C., Ford, D. V. and Lea, N. (2017) 'The other side of the coin: Harm due to the non-use of health-related data', *International Journal of Medical Informatics,* 97, pp. 43-51.
Jones, K. H., McNerney, C. L. and Ford, D. V. (2014) 'Involving consumers in the work of a data linkage research unit', *International Journal of Consumer Studies,* 38(1), pp. 45-51.
Jones, L. A., Nelder, J. R., Fryer, J. M., Alsop, P. H., Geary, M. R., Prince, M. and Cardinal, R. N. (2022b) 'Public opinion on sharing data from health services for clinical and research purposes without explicit consent: an anonymous online survey in the UK', *BMJ Open,* 12(4), pp. e057579.
Kakde, S., Bhopal, R. S. and Jones, C. M. (2012) 'A systematic review on the social context of smokeless tobacco use in the South Asian population: Implications for public health', *Public Health,* 126(8), pp. 635-645.
Kalkman, S., Mostert, M., Gerlinger, C., Johannes, J. M. v. D. and Ghislaine, J. M. W. v. T. (2019a) 'Responsible data sharing in international health research: a systematic review of principles and norms', *BMC Medical Ethics,* 20.
Kalkman, S., Van Delden, J., Banerjee, A., Tyl, B., Mostert, M. and Van Thiel, G. (2019b) 'Patients’ and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence', *Journal of Medical Ethics*, pp. medethics-2019-.
Kamberelis, G. and Dimitriadis, G. (2016) in Denzin, N.K. and Giardina, M.D. (eds.) *Qualitative inquiry and the conservative challenge*: Routledge, pp. 3-30.
Kara, H. (2018) *Research ethics in the real world.* Bristol: Policy Press.
Kara, H. (2020) *Creative research methods: A practical guide.* Policy Press.
Karliner, L. S., Jacobs, E. A., Chen, A. H. and Mutha, S. (2007) 'Do Professional Interpreters Improve Clinical Care for Patients with Limited English Proficiency? A Systematic Review of the Literature', *Health Services Research,* 42(2), pp. 727-754.
Karmegam, D. and Mapillairaju, B. (2020) 'What people share about the COVID-19 outbreak on Twitter? An exploratory analysis', *BMJ Health & Care Informatics,* 27(3), pp. e100133.
Kayrouz, R., Dear, B. F., Karin, E. and Titov, N. (2016) 'Facebook as an effective recruitment strategy for mental health research of hard to reach populations', *Internet Interventions,* 4, pp. 1-10.

Kidder, L. H. and Fine, M. (1987) 'Qualitative and quantitative methods: When stories converge', *New Directions for Program Evaluation,* 1987(35), pp. 57-75.

King, N., Horrocks, C. and Brooks, J. (2018) *Interviews in qualitative research.* sage.

Kirkham, E. J., Crompton, C. J., Iveson, M. H., Beange, I., McIntosh, A. M. and Fletcher-Watson, S. (2021) 'Co-development of a Best Practice Checklist for Mental Health Data Science: A Delphi Study', *Frontiers in Psychiatry,* 12.

Kirwan, G. (2021) 'Superdiversity re-imagined: Applying superdiversity theory to research beyond migration studies', *Current Sociology*, pp. 0011392120983346.

Kosny, A., MacEachen, E., Lifshen, M. and Smith, P. (2014) 'Another Person in the Room: Using Interpreters During Interviews With Immigrant Workers', *Qualitative Health Research,* 24(6), pp. 837-845.

Kumar, B. N., Hargreaves, S., Agyemang, C., James, R. A., Blanchet, K. and Gruer, L. (2021) 'Reducing the impact of the coronavirus on disadvantaged migrants and ethnic minorities', *European Journal of Public Health,* 31(Supplement_4), pp. iv9-iv13.

Kurath, M. and Gisler, P. (2009) 'Informing, involving or engaging? Science communication, in the ages of atom-, bio- and nanotechnology', *Public Understanding of Science,* 18(5), pp. 559-573.

Laird, Y., Manner, J., Baldwin, L., Hunter, R., Mcateer, J., Rodgers, S., Williamson, C. and Jepson, R. (2020) 'Stakeholders' experiences of the public health research process: time to change the system?', *Health Research Policy and Systems,* 18(1).

Lalani, M., Baines, R., Bryce, M., Marshall, M., Mead, S., Barasi, S., Archer, J. and Regan de Bere, S. (2019) 'Patient and public involvement in medical performance processes: A systematic review', *Health expectations : an international journal of public participation in health care and health policy,* 22(2), pp. 149-161.

Largent, E. A. and Fernandez Lynch, H. (2017) 'Paying Research Participants: Regulatory Uncertainty, Conceptual Confusion, and a Path Forward', *Yale J Health Policy Law Ethics,* 17(1), pp. 61-141.

Lauzon-Schnittka, J., Audette-Chapdelaine, S., Boutin, D., Wilhelmy, C., Auger, A.-M. and Brodeur, M. (2022) 'The experience of patient partners in research: a qualitative systematic review and thematic synthesis', *Research Involvement and Engagement,* 8(1).

Lawrence, N. R. and Bradley, S. H. (2018) 'Big data and the NHS - we have the -technology, but we need patient and professional engagement', *Future healthcare journal,* 5(3), pp. 229-230.

Lee, N. M. and Vandyke, M. S. (2015) 'Set It and Forget It: The One-Way Use of Social Media by Government Agencies Communicating Science', *Science Communication,* 37(4), pp. 533-541.

Lee, R. G. and Garvin, T. (2003) 'Moving from information transfer to information exchange in health and health care', *Social Science & Medicine,* 56(3), pp. 449-464.

Leinonen, T. and Tandefelt, M. (2007) 'Evidence of language loss in progress? Mother-tongue proficiency among students in Finland and Sweden'.

Levac, D., Colquhoun, H. and O'Brien, K. K. (2010) 'Scoping studies: advancing the methodology', *Implementation Science,* 5(1), pp. 69.

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, C. J., Gülmezoglu, M., Noyes, J., Booth, A., Garside, R. and Rashidian, A. (2015) 'Using Qualitative Evidence in Decision Making for Health and Social Interventions: An Approach to Assess Confidence in Findings from Qualitative Evidence Syntheses (GRADE-CERQual)', *PLOS Medicine,* 12(10), pp. e1001895.

Lincoln, Y. S. and Guba, E. G. (1985) *Naturalistic inquiry.* sage.

Lincoln, Y. S. and Tierney, W. G. (2004) 'Qualitative Research and Institutional Review Boards', *Qualitative Inquiry,* 10(2), pp. 219-234.

Lipworth, W., Mason, P. H., Kerridge, I. and Ioannidis, J. P. A. (2017) 'Ethics and Epistemology in Big Data Research', *Journal of Bioethical Inquiry,* 14(4), pp. 489-500.

Lo Iacono, V., Symonds, P. and Brown, D. H. K. (2016) 'Skype as a Tool for Qualitative Research Interviews', *Sociological Research Online,* 21(2), pp. 103-117.

Lobe, B., Morgan, D. and Hoffman, K. A. (2020) 'Qualitative Data Collection in an Era of Social Distancing', *International Journal of Qualitative Methods,* 19, pp. 160940692093787.

Lomas, N. (2021) *Google confirms it’s pulling the plug on Streams, its UK clinician support app*. Available at: https://techcrunch.com/2021/08/26/google-confirms-its-pulling-the-plug-on-streams-its-uk-clinician-support-app/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cudWt0ZWNoLm5ld3MvYWkvZ29vZ2xlLWxhd3N1aXQtZGVlcG1pbmQtbmhzLWRhdGEtMjAyMjA1MTc&guce_referrer_sig=AQAAAIvu3A0K2fY8_eWlmx8R52bNFkO2LR0JqfLzXFIG5QZVCZEA1f7bMqPuWylD40WqOMaVq7D1qOqaitplkZmXtepuTn2VNvckKVYawMyo737ZfqxzpJpwCC8s3gYLJ_0oKzdxik-x7877j_6MX57uM4P3eqhDEbhRiI3feWjnQDNp (Accessed: 19 February 2023).

López-Goñi, I. and Sánchez-Angulo, M. (2018) 'Social networks as a tool for science communication and public engagement: focus on Twitter', *FEMS Microbiology Letters,* 365(2).

Luna Puerta, L., Kendall, W., Davies, B., Day, S. and Ward, H. (2020) 'The reported impact of public involvement in biobanks: A scoping review', *Health Expectations,* 23(4), pp. 759-788.

Luna Puerta, L. and Smith, H. E. (2020) 'The “PPI Hawker”: an innovative method for patient and public involvement (PPI) in health research', *Research Involvement and Engagement,* 6(1).

Lynch, M. (2000) 'Against Reflexivity as an Academic Virtue and Source of Privileged Knowledge', *Theory, Culture & Society,* 17(3), pp. 26-54.

Macaulay, A. C. (2016) 'Participatory research: What is the history? Has the purpose changed?', *Family Practice*, pp. cmw117.

Madden, M. and Speed, E. (2017) 'Beware Zombies and Unicorns: Toward Critical Patient and Public Involvement in Health Research in a Neoliberal Context', *Frontiers in Sociology,* 2.

Malterud, K., Siersma, V. D. and Guassora, A. D. (2015) 'Sample Size in Qualitative Interview Studies: Guided by Information Power', *Qualitative Health Research,* 26(13), pp. 1753-1760.

Manafo, E., Petermann, L., Mason-Lai, P. and Vandall-Walker, V. (2018) 'Patient engagement in Canada: a scoping review of the 'how' and 'what' of patient engagement in health research', *Health Res Policy Syst,* 16(1), pp. 5.

Manhas, K. P., Page, S., Dodd, S. X., Letourneau, N., Ambrose, A., Cui, X. and Tough, S. C. (2015) 'Parent Perspectives on Privacy and Governance for a Pediatric Repository of Non-Biological, Research Data', *Journal of Empirical Research on Human Research Ethics,* 10(1), pp. 88-99.

Manikandan, M., Foley, K., Gough, J., Harrington, S., Wall, É., Weldon, F., Ryan, J. M., Kerr, C., Walsh, A. and Fortune, J. (2022) 'Public and Patient Involvement in Doctoral Research During the COVID-19 Pandemic: Reflections on the Process, Challenges, Impact and Experiences From the Perspectives of Adults With Cerebral Palsy and the Doctoral Researcher', *Frontiers in Rehabilitation Sciences,* 3.

Manrique de Lara, A. and Peláez-Ballestas, I. (2020) 'Big data and data processing in rheumatology: bioethical perspectives', *Clinical Rheumatology,* 39(4), pp. 1007-1014.

Mapedzahama, V. and Kwansah-Aidoo, K. (2016) 'Black on Black: Insider positionality and the Black African migrant research experience in Australia', *Insider Research on Migration and Mobility*: Routledge, pp. 169-186.

Masterson, D., Areskoug Josefsson, K., Robert, G., Nylander, E. and Kjellström, S. (2022) 'Mapping definitions of co-production and co-design in health and social care: A systematic scoping review providing lessons for the future', *Health Expectations*.

McCormack, B. and McCance, T. (2016) *Person-centred practice in nursing and health care: theory and practice.* John Wiley & Sons.

Mcgrath, C., Palmgren, P. J. and Liljedahl, M. (2019) 'Twelve tips for conducting qualitative research interviews', *Medical Teacher,* 41(9), pp. 1002-1006.

McGrath, J. (2016) *Co-production: an inconvenient truth?* Available at: https://www.kingsfund.org.uk/blog/2016/10/co-production-inconvenient-truth (Accessed: 28 January 2021).

Mehta, N. and Pandit, A. (2018) 'Concurrence of big data analytics and healthcare: A systematic review', *International Journal of Medical Informatics,* 114, pp. 57-65.

Melis, G., Sala, E. and Zaccaria, D. (2022) 'Remote recruiting and video-interviewing older people: a research note on a qualitative case study carried out in the first Covid-19 Red Zone in Europe', *International Journal of Social Research Methodology,* 25(4), pp. 477-482.

Merry, L., Clausen, C., Gagnon, A. J., Carnevale, F., Jeannotte, J., Saucier, J.-F. and Oxman-Martinez, J. (2011) 'Improving Qualitative Interviews With Newly Arrived Migrant Women', *Qualitative Health Research,* 21(7), pp. 976-986.

Metcalfe, J. (2019) 'Comparing science communication theory with practice: An assessment and critique using Australian data', *Public Understanding of Science,* 28(4), pp. 382-400.

Michie, S., Van Stralen, M. M. and West, R. (2011) 'The behaviour change wheel: A new method for characterising and designing behaviour change interventions', *Implementation Science,* 6(1), pp. 42.

Michie, S., West, R., Campbell, R., Brown, J. and Gainforth, H. (2014) *ABC of behaviour change theories.* Silverback publishing.

Miller, S. (2001) 'Public understanding of science at the crossroads', *Public Understanding of Science,* 10(1), pp. 115-120.

Milne, R. and Brayne, C. (2020) 'We need to think about data governance for dementia research in a digital era', *Alzheimer's Research and Therapy,* 12(1).

Milne, R., Sorbie, A. and Dixon-Woods, M. (2021) 'What can data trusts for health research learn from participatory governance in biobanks?', *Journal of Medical Ethics*.

Mockford, C., Staniszewska, S., Griffiths, F. and Herron-Marx, S. (2012) 'The impact of patient and public involvement on UK NHS health care: a systematic review', *International Journal for Quality in Health Care,* 24(1), pp. 28-38.

Moffat, K., Lacey, J., Zhang, A. and Leipold, S. (2016) 'The social licence to operate: a critical review', *Forestry: An International Journal of Forest Research,* 89(5), pp. 477-488.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (2009) 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement', *PLoS Medicine,* 6(7), pp. e1000097.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P. and Stewart, L. A. (2015) 'Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement', *Systematic Reviews,* 4(1), pp. 1.

Monaghan, L. F., O'Dwyer, M. and Gabe, J. (2013) 'Seeking university Research Ethics Committee approval: the emotional vicissitudes of a 'rationalised' process', *International Journal of Social Research Methodology,* 16(1), pp. 65-80.

Moody, G., Cannings-John, R., Hood, K. and Robling, M. (2022) 'Attitudes towards the collection and linkage of maltreatment data for research: A qualitative study', *International Journal of Population Data Science,* 7(1).

MORI, I. (2017) 'The One-Way Mirror: Public attitudes to commercial access to health data'.

Morrow, E., Boaz, A. and Sally, B. F. (2011) 'Seldom-Heard Groups', *Handbook of service user involvement in nursing and healthcare research*, pp. 106-119.

Morse, J. M. (2000) 'Determining Sample Size', *Qualitative Health Research,* 10(1), pp. 3-5.

Morse, J. M. (2010) '"Cherry Picking": Writing From Thin Data', *Qualitative Health Research,* 20(1), pp. 3-3.

Mortelmans, D. (2019) 'Analyzing qualitative data using NVivo', *The Palgrave handbook of methods for media policy research*: Springer, pp. 435-450.

Mourby, M., Doidge, J., Jones, K. H., Aidinlis, S., Smith, H., Bell, J., Gilbert, R., Dutey-Magni, P. and Kaye, J. (2019) 'Health data linkage for UK public interest research: Key obstacles and solutions', *International Journal of Population Data Science,* 4(1).

Mouton Dorey, C., Baumann, H. and Biller-Andorno, N. (2018) 'Patient data and patient rights: Swiss healthcare stakeholders' ethical awareness regarding large patient data sets - a qualitative study', *BMC medical ethics,* 19(1), pp. 20-20.

Mueller, A., Johnston, M., Bligh, D. and Wilkinson, J. (2002) 'Joining Mind Mapping and Care Planning to Enhance Student Critical Thinking and Achieve Holistic Nursing Care', *International Journal of Nursing Terminologies and Classifications,* 13(1), pp. 24-27.

Muller, S. H. A., Kalkman, S., Van Thiel, G. J. M. W., Mostert, M. and Van Delden, J. J. M. (2021) 'The social licence for data-intensive health research: towards co-creation, public value and trust', *BMC Medical Ethics,* 22(1).

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., Mcarthur, A. and Aromataris, E. (2018) 'Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach', *BMC Medical Research Methodology,* 18(1).

Murtagh, M. J., Blell, M. T., Butters, O. W., Cowley, L., Dove, E. S., Goodman, A., Griggs, R. L., Hall, A., Hallowell, N., Kumari, M., Mangino, M., Maughan, B., Mills, M. C., Minion, J. T., Murphy, T., Prior, G., Suderman, M., Ring, S. M., Rogers, N. T., Roberts, S. J., Van Der Straeten, C., Viney, W., Wiltshire, D., Wong, A., Walker, N. and Burton, P. R. (2018) 'Better governance, better access: Practising responsible data sharing in the METADAC governance infrastructure', *Human Genomics,* 12(1).

National Co-ordinating Centre for Public Engagement (2021) *What is public engagement?* Available at: https://www.publicengagement.ac.uk/about-engagement/what-public-engagement (Accessed: 17 November 2021).

Nelson, E. and Burns, F. (2020) 'Impact through Engagement. Co-production of administrative data research andthe approach of the Administrative Data Research Centre Northern Ireland', *International Journal of Population Data Science,* 5(3).

Neuberger, J. and Tallis, R. (1999) 'Do we need a new word for patients?   Let's do away with "patients"   Commentary: Leave well alone', *BMJ,* 318(7200), pp. 1756-1758.

Newburn, M., Scanlon, M., Plachcinski, R. and Macfarlane, A. J. (2020) 'Involving service users in Births and their outcomes: a retrospective birth cohort data linkage study analysing daily, weekly and yearly cycles and their implications for the NHS', *International Journal of Population Data Science,* 5(3).

Newington, L., Alexander, C. M., Kirby, P., Saggu, R. K. and Wells, M. (2022) 'Reflections on contributing to health research: A qualitative interview study with research participants and patient advisors', *PLOS ONE,* 17(12), pp. e0279354.

Nguyen, T., H., Cheah, P. Y. and Chambers, M. (2019) 'Identifying 'hard-to-reach' groups and strategies to engage them in biomedical research: perspectives from engagement practitioners in Southeast Asia', *Wellcome open research,* 4.

NHS (2019) *Research Ethics Committee – Standard Operating Procedures*.

NHS (2020) *UK Policy Framework for Health and Social Care Research*.

NHS Digital (2021) *General Practice Data for Planning and Research (GPDPR)*. Available at: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research#top (Accessed: 17 November 2021).

NHS Digital (2022) *About the GPDPR programme*. Available at: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research/about-the-gpdpr-programme.

NHS England (2017) *Patient and public participation policy*. Available at: https://www.england.nhs.uk/wp-content/uploads/2017/04/ppp-policy.pdf (Accessed: 22 January 2021).

NIHR (2019) *National Standards for Public Involvement*.

NIHR (2020) *Improving inclusion of under-served groups in clinical research: Guidance from INCLUDE project*. Available at: https://www.nihr.ac.uk/documents/improving-inclusion-of-under-served-groups-in-clinical-research-guidance-from-include-project/25435 (Accessed: 5 November 2020).

NIHR (2021a) *Briefing notes for researchers - public involvement in NHS, health and social care research*. Available at: https://www.nihr.ac.uk/documents/briefing-notes-for-researchers-public-involvement-in-nhs-health-and-social-care-research/27371 (Accessed: 14 July 2022).

NIHR (2021b) *Get Involved*. Available at: https://arc-nwc.nihr.ac.uk/get-involved/ (Accessed: 20 January 2021).

NIHR (2021c) *Next steps for partnership working with patients and the public: Engagement report*. Available at: https://www.nihr.ac.uk/documents/next-steps-for-partnership-working-with-patients-and-the-public-engagement-report/28566 (Accessed: 28 July 2022).
NIHR (2023) *22/136 Public acceptance of the use of administrative datasets for Public Health Research*. Available at: https://www.nihr.ac.uk/funding/22136-public-acceptance-of-the-use-of-administrative-datasets-for-public-health-research/31756 (Accessed: 6 April 2023).
NMC (2018) *The Code: Professional standards of practice and behaviour for nurses, midwives and nursing associates*.
Nunn, J. S., Gwynne, K., Gray, S. and Lacaze, P. (2021) 'Involving people affected by a rare condition in shaping future genomic research', *Research Involvement and Engagement,* 7(1).
NVivo (2022) *NCapture*. Available at: https://support.qsrinternational.com/nvivo/s/article/NVR1Win-NCapture (Accessed: 10 October 2022).
NVivo (2023) *About NVivo*. Available at: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/about/nvivo (Accessed: 25 January 2023).
O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A. and Cook, D. A. (2014) 'Standards for Reporting Qualitative Research', *Academic Medicine,* 89(9), pp. 1245-1251.
O'Doherty, K. C., Burgess, M. M., Edwards, K., Gallagher, R. P., Hawkins, A. K., Kaye, J., Mccaffrey, V. and Winickoff, D. E. (2011) 'From consent to institutions: Designing adaptive governance for genomic biobanks', *Social Science & Medicine,* 73(3), pp. 367-374.
O'Doherty, K. C., Shabani, M., Dove, E. S., Bentzen, H. B., Borry, P., Burgess, M. M., Chalmers, D., De Vries, J., Eckstein, L., Fullerton, S. M., Juengst, E., Kato, K., Kaye, J., Knoppers, B. M., Koenig, B. A., Manson, S. M., McGrail, K. M., McGuire, A. L., Meslin, E. M., Nicol, D., Prainsack, B., Terry, S. F., Thorogood, A. and Burke, W. (2021) 'Toward better governance of human genomic data', *Nature Genetics,* 53(1), pp. 2-8.
O'Dowd, A. (2021) 'New GP patient data extraction scheme raises concern', *BMJ,* 373, pp. n1389.
Ocloo, J. and Matthews, R. (2016) 'From tokenism to empowerment: progressing patient and public involvement in healthcare improvement', *BMJ Quality & Safety,* 25(8), pp. 626-632.
OED (1990) *Involvement*. Available at: https://www.oed.com/view/Entry/99208?redirectedFrom=Involvement#eid (Accessed: 12 November 2020).
Ohno-Machado, L., Agha, Z., Bell, D. S., Dahm, L., Day, M. E., Doctor, J. N., Gabriel, D., Kahlon, M. K., Kim, K. K., Hogarth, M., Matheny, M. E., Meeker, D. and Nebeker, J. R. (2014) 'pSCANNER: Patient-centered scalable national network for effectiveness research', *Journal of the American Medical Informatics Association,* 21(4), pp. 621-626.
Omar Muhammad, I. and Roobol, M. J. and Ribal, M. J. and Abbott, T. and Paul-Michael, A. and Araujo, S. and Asiimwe, A. and Auffray, C. and Balaur, I. and Beyer, K. and Bernini, C. and Bjartell, A. and Briganti, A. and John-Edward, B.-R. and Campi, R. and Cavelaars, M. and De Meulder, B. and Devecseri, Z. and Voss, M. D. and Dimitropoulos, K. and Evans-Axelsson, S. and Franks, B. and Fullwood, L. and Horgan, D. and Smith, E. J. and Kiran, A. and Kivinummi, K. and Lambrecht, M. and Lancet, D. and Lindgren, P. and MacLennan, S. and MacLennan, S. and Nogueira, M. M. and Moen, F. and Moinat, M. and Papineni, K. and Reich, C. and Reiche, K. and Rogiers, S. and Sartini, C. and Kees, v. B. and Femke, v. D. and Mieke, V. H. and Van Poppel, H. and Zong, J. and N'Dow, J. and Andersson, E. and Arala, H. and Auvinen, A. and Bangma, C. and Burke, D. and Cardone, A. and Casariego, J. and Cuperus, G. and Dabestani, S. and Esperto, F. and Fossati, N. and Fridhammar, A. and Gandaglia, G. and Tandefelt Delila, G. and Horn, F. and Huber, J. and Hugosson, J. and Huisman, H. and Josefsson, A. and Kilkku, O. and Kreuz, M. and Lardas, M. and Lawson, J. and Lefresne, F. and Lejeune, S. and Longden-Chapman, E. and McVie, G. and Moris, L. and Mottet, N. and Murtola, T. and Nicholls, C. and Pang, K. H. and Pascoe, K. and Picozzi, M. and Plass, K. and Pohjanjousi, P. and Reaney, M. and Remmers, S. and Robinson, P. and Schalken, J. and Schravendeel, M. and Seisen, T. and Servan, A. and Shiranov, K. and Snijder, R. and Steinbeisser, C. and Taibi, N. and Talala, K. and Tilki, D. and Van

den Broeck, T. and Vassilev, Z. and Voima, O. and Vradi, E. and Waldeck, R. and Ward, W. and Peter-Paul, W. and Wirth, M. and Wolfinger, R. and Kermani, N. Z. (2020) 'Introducing PIONEER: a project to harness big data in prostate cancer research', *Nature Reviews. Urology,* 17(6), pp. 351-362.
OpenSAFELY (2021) *OpenSAFELY: About Us*. Available at: https://www.opensafely.org/about/ (Accessed: 22 June 2021).
Oxford, E. (2022) 'Hundreds of patient data breaches are left unpunished', *BMJ,* 377, pp. o1126.
Paprica, P. A., De Melo, M. N. and Schull, M. J. (2019) 'Social licence and the general public's attitudes toward research based on linked administrative health data: a qualitative study', *CMAJ Open,* 7(1), pp. E40-E46.
Paprica, P. A., Sutherland, E., Smith, A., Brudno, M., Cartagena, R. G., Crichlow, M., Courtney, B. K., Loken, C., McGrail, K. M., Ryan, A., Schull, M. J., Thorogood, A., Virtanen, C. and Yang, K. (2020) 'Essential requirements for establishing and operating data trusts: practical guidance co-developed by representatives from fifteen canadian organizations and initiatives', *International journal of population data science,* 5(1), pp. 1353.
Parkinson, J. (2004) 'Hearing Voices: Negotiating Representation Claims in Public Deliberation', *The British Journal of Politics and International Relations,* 6(3), pp. 370-388.
Patel, R., Irving, J., Brinn, A., Broadbent, M., Shetty, H., Pritchard, M., Downs, J., Stewart, R., Harland, R. and McGuire, P. (2021) 'Impact of the COVID-19 pandemic on remote mental healthcare and prescribing in psychiatry: an electronic health record study', *BMJ Open,* 11(3).
Patients Included (2020) *What is the purpose of the Patients Included journal charter?* Available at: https://patientsincluded.org/journals/ (Accessed: 10 January 2021).
Patton, M. Q. (2014) *Qualitative research & evaluation methods: Integrating theory and practice.* Sage publications.
Paucsik, M., Leys, C., Marais, G., Baeyens, C. and Shankland, R. (2022) 'Self-compassion and savouring buffer the impact of the first year of the COVID-19 on PhD students' mental health', *Stress and Health*.
Paul, J., Davidson, R., Johnstone, C., Loong, M., Matecsa, J., Guttmann, A. and Schull, M. J. (2020) 'Public engagement can change your research, but how can it change your research institution? ICES Case Study', *International Journal of Population Data Science,* 5(3).
Pavlenko, E., Strech, D. and Langhof, H. (2020) 'Implementation of data access and use procedures in clinical data warehouses. A systematic review of literature and publicly available policies', *BMC Medical Informatics and Decision Making,* 20(1).
PEDRI (2023) *Public Engagement in Data Research Initiative*. Available at: https://www.pedri.org.uk/ (Accessed: 27 March 2023).
Pfadenhauer, M. (2018) 'Introduction: The reality of social constructivism Introductory remarks', *Social Constructivism as Paradigm?*: Routledge, pp. 1-18.
Phillimore, J. (2011) 'Approaches to health provision in the age of super-diversity: accessing the NHS in Britain's most diverse city', *Critical Social Policy,* 31(1), pp. 5-29.
Phillips, E. (2022) *Why we want to recruit data sceptics*. Available at: https://digital.nhs.uk/blog/data-points-blog/2022/why-we-want-to-recruit-our-data-sceptics (Accessed: 20 February 2023).
Pii, K. H., Schou, L. H., Piil, K. and Jarden, M. (2019) 'Current trends in patient and public involvement in cancer research: A systematic review', *Health expectations : an international journal of public participation in health care and health policy,* 22(1), pp. 3-20.
Plamping, D. and Delamothe, T. (1991) 'The citizen's charter and the NHS.', *BMJ,* 303(6796), pp. 203-204.
Polacsek, M., Boardman, G. and McCann, T. V. (2017) 'Paying patient and caregiver research participants: putting theory into practice', *Journal of Advanced Nursing,* 73(4), pp. 847-856.
Portland Communications (2021) *Analysis of UK reporting on health data*: Understanding Patient Data. Available at: https://understandingpatientdata.org.uk/analysis-uk-media-reporting-health-data-stories (Accessed: 24 February 2023).

Pötzschke, S. (2022) 'Using the innovative to improve the established: The employment of social networking sites as recruitment tools in migrant surveys', *International Migration,* 60(2), pp. 261-265.
Pötzschke, S. and Braun, M. (2017) 'Migrant Sampling Using Facebook Advertisements:A Case Study of Polish Migrants in Four European Countries', *Social Science Computer Review,* 35(5), pp. 633-653.
Prajapati, R. and Liebling, H. (2022) 'Accessing Mental Health Services: a Systematic Review and Meta-ethnography of the Experiences of South Asian Service Users in the UK', *Journal of Racial and Ethnic Health Disparities,* 9(2), pp. 598-619.
Priest, S., Goodwin, J. and Dahlstrom, M. F. (2018) *Ethics and practice in science communication.* University of Chicago Press.
Prinjha, S., Miah, N., Ali, E. and Farmer, A. (2020) 'Including ‘seldom heard’ views in research: opportunities, challenges and recommendations from focus groups with British South Asian people with type 2 diabetes', *BMC Medical Research Methodology,* 20(1).
Raghupathi, W. and Raghupathi, V. (2014) 'Big data analytics in healthcare: promise and potential', *Health Information Science and Systems,* 2(1), pp. 3.
Ramesh, R. (2014) *NHS patient data to be made available for sale to drug and insurance firms*: The Guardian. Available at: https://www.theguardian.com/society/2014/jan/19/nhs-patient-data-available-companies-buy (Accessed: 19 February 2023).
Ranasinghe, C. D., Ranasinghe, P., Jayawardena, R. and Misra, A. (2013) 'Physical activity patterns among South-Asian adults: a systematic review', *International Journal of Behavioral Nutrition and Physical Activity,* 10(1), pp. 116.
Rashid, M. A., Mckechnie, D. and Gill, D. (2018) 'What advice is given to newly qualified doctors on Twitter? An analysis of #TipsForNewDocs tweets', *Medical Education,* 52(7), pp. 747-756.
Rayment, J., Lanlehin, R., Mccourt, C. and Husain, S. M. (2017) 'Involving seldom-heard groups in a PPI process to inform the design of a proposed trial on the use of probiotics to prevent preterm birth: a case study', *Research Involvement and Engagement,* 3(1).
Rees, L., Sherwood, M. and Shields, N. (2022) 'Consumer engagement in doctoral research – what difference does it make?', *Spinal Cord*.
Reflecting on the online movement – how social media is spreading the message of Data Saves Lives (2021). Twitter Moment. Available at: https://twitter.com/i/events/1464220352674996242 (Accessed: 26 November 2021).
Renedo, A., Marston, C. A., Spyridonidis, D. and Barlow, J. (2015) 'Patient and Public Involvement in Healthcare Quality Improvement: How organizations can help patients and professionals to collaborate', *Public Management Review,* 17(1), pp. 17-34.
Revell, T. (2017) *Google DeepMind’s NHS data deal 'failed to comply' with law*. New Scientist. Available at: https://institutions.newscientist.com/article/2139395-google-deepminds-nhs-data-deal-failed-to-comply-with-law/.
Rhodes, P., Nocon, A., Booth, M., Chowdrey, M. Y., Fabian, A., Lambert, N., Mohammed, F. and Walgrove, T. (2002) 'A service users’ research advisory group from the perspectives of both service users and researchers', *Health & Social Care in the Community,* 10(5), pp. 402-409.
Richardson, J., Grose, J., Nelmes, P., Parra, G. and Linares, M. (2016) 'Tweet if you want to be sustainable: a thematic analysis of a Twitter chat to discuss sustainability in nurse education', *Journal of Advanced Nursing,* 72(5), pp. 1086-1096.
Ritchie, J., Spencer, L., Bryman, A. and Burgess, R. G. 1994. Analysing qualitative data. Routledge, London.
Rohwer, A., Pfadenhauer, L., Burns, J., Brereton, L., Gerhardus, A., Booth, A., Oortwijn, W. and Rehfuess, E. (2017) 'Series: Clinical Epidemiology in South Africa. Paper 3: Logic models help make sense of complexity in systematic reviews and health technology assessments', *Journal of Clinical Epidemiology,* 83, pp. 37-47.
Romaioli, D. and McNamee, S. (2021) '(Mis)constructing social construction: Answering the critiques', *Theory & Psychology,* 31(3), pp. 315-334.

Romsland, G. I., Milosavljevic, K. L. and Andreassen, T. A. (2019) 'Facilitating non-tokenistic user involvement in research', *Research Involvement and Engagement,* 5(1).
Rosenberg, E., Seller, R. and Leanza, Y. (2008) 'Through interpreters' eyes: Comparing roles of professional and family interpreters', *Patient Education and Counseling,* 70(1), pp. 87-93.
Rowe, R. K., Carroll, S. R., Healy, C., Rodriguez-Lonebear, D. and Walker, J. D. (2021) 'The SEEDS of indigenous population health data linkage', *International Journal of Population Data Science,* 6(1).
Royal Free London NHS Foundation Trust (2017) *New app helping to improve patient care*. Available at: https://www.royalfree.nhs.uk/news-media/news/new-app-helping-to-improve-patient-care/ (Accessed: 19 February 2023).
Rubin, H. J. and Rubin, I. S. (2005) *Qualitative interviewing: The art of hearing data.* Sage.
Rubinstein, H., Marcu, A., Yardley, L. and Michie, S. (2015) 'Public preferences for vaccination and antiviral medicines under different pandemic flu outbreak scenarios', *BMC Public Health,* 15(1), pp. 190.
Rzepnikowska, A. (2018) 'Racism and xenophobia experienced by Polish migrants in the UK before and after Brexit vote', *Journal of Ethnic and Migration Studies*, pp. 1-17.
Salamońska, J., Lesińska, M. and Kloc-Nowak, W. (2021) 'Polish Migrants in Ireland and Their Political (Dis) engagement in Transnational Space', *Central and Eastern European Migration Review,* 10(2), pp. 49-69.
Sandelowski, M. (1995) 'Sample size in qualitative research', *Research in nursing & health,* 18(2), pp. 179-183.
Schwandt, T. A. (2000) 'Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism', *Handbook of qualitative research*: SAGE Publishing, pp. 189-213.
Scott, P., Emerson, K. and Henderson-Reay, T. (2021) 'Data saves lives', *BMJ,* 374, pp. n1694.
Seers, K. (2012) 'Qualitative data analysis', *Evidence Based Nursing,* 15(1), pp. 2-2.
Seethaler, S., Evans, J. H., Gere, C. and Rajagopalan, R. M. (2019) 'Science, Values, and Science Communication: Competencies for Pushing Beyond the Deficit Model', *Science Communication,* 41(3), pp. 378-388.
Shabani, M., Bezuidenhout, L. and Borry, P. (2014) 'Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review', *Expert Review of Molecular Diagnostics,* 14(8), pp. 1053-1065.
Shah, N., Coathup, V., Teare, H., Forgie, I., Giordano, G. N., Hansen, T. H., Groeneveld, L., Hudson, M., Pearson, E., Ruetten, H. and Kaye, J. (2019) 'Motivations for data sharing-views of research participants from four European countries: A DIRECT study', *Eur J Hum Genet,* 27(5), pp. 721-729.
Sharma, M. (2018) ''Can the patient speak?': postcolonialism and patient involvement in undergraduate and postgraduate medical education', *Medical Education,* 52(5), pp. 471-479.
Sharp, C. and Randhawa, G. (2017) 'Conducting cross-cultural interviews and focus groups concerning healthcare with Polish migrants in the UK – Lessons from a study on organ donation', *Polish Annals of Medicine,* 24(1), pp. 13-18.
Shaun, L. (2021) 'Misinformation on NHS data sharing could scupper vital health research, warn experts', *The Independent*. Available at: https://www.independent.co.uk/news/health/nhs-data-sharing-research-health-b1862658.html (Accessed: 29 July 2022).
Shaw, D. and Erren, T. (2021) 'If data save lives, consent saves data and data donation', *BMJ,* 374, pp. n2069.
Shaw, J. A., Sethi, N. and Cassel, C. K. (2020) 'Social license for the use of big data in the COVID-19 era', *NPJ Digital Medicine,* 3(1).
Silva, D. S., Smith, M. J. and Upshur, R. E. G. (2013) 'Disadvantaging the Disadvantaged: When Public Health Policies and Practices Negatively Affect Marginalized Populations', *Canadian Journal of Public Health,* 104(5), pp. e410-e412.

Simis, M. J., Madden, H., Cacciatore, M. A. and Yeo, S. K. (2016) 'The lure of rationality: Why does the deficit model persist in science communication?', *Public Understanding of Science,* 25(4), pp. 400-414.
Simister, G. (2022) *Google faces lawsuit over DeepMind's NHS data use*. Available at: https://www.uktech.news/ai/google-lawsuit-deepmind-nhs-data-20220517 (Accessed: 19 February 2023).
Simmons, P., Hawley, C. J., Gale, T. M. and Sivakumaran, T. (2010) 'Service user, patient, client, user or survivor: describing recipients of mental health services', *The Psychiatrist,* 34(1), pp. 20-23.
Simpson, E. L., Bryant, L. D., Ruddy, R. and Atkin, K. (2006) 'Developing partnerships for research: training workshops for mental health service users, carers and workers', *Psychiatric Bulletin,* 30(2), pp. 67-69.
Sipes, J. B. A., Roberts, L. D. and Mullan, B. (2022) 'Voice-only Skype for use in researching sensitive topics: a research note', *Qualitative Research in Psychology,* 19(1), pp. 204-220.
Skovgaard, L. L., Wadmann, S. and Hoeyer, K. (2019) 'A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good', *Health Policy,* 123(6), pp. 564-571.
Sleigh, J. and Vayena, E. (2021) 'Public engagement with health data governance: the role of visuality', *Humanities and Social Sciences Communications,* 8(1).
Smith, J. A. and Fieldsend, M. (2021) *Interpretative phenomenological analysis.* American Psychological Association.
Snow, M. E., Tweedie, K. and Pederson, A. (2018) 'Heard and valued: the development of a model to meaningfully engage marginalized populations in health services planning', *BMC Health Services Research,* 18(1).
Sofolahan-Oladeinde, Y., Newhouse, R. P., Lavallee, D. C., Huang, J. C. and Mullins, C. D. (2017) 'Early assessment of the 10-step patient engagement framework for patient-centred outcomes research studies: the first three steps', *Family Practice,* 34(3), pp. 272-277.
Sotiriadou, P., Brouwers, J. and Le, T.-A. (2014) 'Choosing a qualitative data analysis tool: a comparison of NVivo and Leximancer', *Annals of Leisure Research,* 17(2), pp. 218-234.
Spencer, K., Sanders, C., Whitley, E. A., Lund, D., Kaye, J. and Dixon, W. G. (2016) 'Patient Perspectives on Sharing Anonymized Personal Health Data Using a Digital System for Dynamic Consent and Research Feedback: A Qualitative Study', *Journal of Medical Internet Research,* 18(4).
Squires, A. (2008) 'Language barriers and qualitative nursing research: methodological considerations', *International Nursing Review,* 55(3), pp. 265-273.
Staniszewska, S., Brett, J., Simera, I., Seers, K., Mockford, C., Goodlad, S., Altman, D. G., Moher, D., Barber, R., Denegri, S., Entwistle, A., Littlejohns, P., Morris, C., Suleman, R., Thomas, V. and Tysall, C. (2017a) 'GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research', *BMJ*, pp. j3453.
Staniszewska, S., Brett, J., Simera, I., Seers, K., Mockford, C., Goodlad, S., Altman, D. G., Moher, D., Barber, R., Denegri, S., Entwistle, A., Littlejohns, P., Morris, C., Suleman, R., Thomas, V. and Tysall, C. (2017b) 'GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research', *Research Involvement and Engagement,* 3(1).
Staniszewska, S., Jones, N., Newburn, M. and Marshall, S. (2007) 'User involvement in the development of a research bid: barriers, enablers and impacts', *Health Expectations,* 10(2), pp. 173-183.
Starbuck, L., Walker, K., Welch, J., Rigby, E. and Hagell, A. (2020) 'Innovative ways of engaging young people whose voices are less heard', in Brady, L.-M. (ed.) *Embedding Young People's Participation in Health Services. New Approaches*. Bristol: Policy Press, pp. 131-153.
Steel, R. (2005a) 'Actively Involving Marginalized and Vulnerable People in Research', in Lowes, L. and Hulatt, I. (eds.) *Involving Service Users in Health and Social Care Research*. Abingdon: Routledge.
Steel, R. (2005b) 'Actively Involving Marginalized and Vulnerble People in Research', in Lowes, L. and Hulatt, I. (eds.) *Involving Service Users in Health and Social Care Research*. Abingdon: Routledge.

Stewart, D. (2022) *A History of Patient Involvement (particularly in cancer research)*. Available at: https://involveddotblog.wordpress.com/a-history-of-patient-involvement-particularly-in-cancer-research/ (Accessed: 1 February 2023).

Stockdale, J., Cassell, J. and Ford, E. (2019) '“Giving something back”: A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland', *Wellcome Open Research,* 3, pp. 6.

Stocking, B. (1991) 'Patient's charter.', *BMJ,* 303(6811), pp. 1148-1149.

Strauss, A. and Corbin, J. M. (1997) *Grounded theory in practice.* Sage.

Strauss, A. L. (1990) *Basics of qualitative research : grounded theory procedures and techniques.* Newbury Park, Calif: Sage.

Su, L. Y.-F., Scheufele, D. A., Bell, L., Brossard, D. and Xenos, M. A. (2017) 'Information-Sharing and Community-Building: Exploring the Use of Twitter in Science Public Relations', *Science Communication,* 39(5), pp. 569-597.

Suleyman, M. and King, D. (2017) *The Information Commissioner, the Royal Free, and what we’ve learned*. Available at: https://www.deepmind.com/blog/the-information-commissioner-the-royal-free-and-what-weve-learned (Accessed: 19 February 2023).

Sunstein, C. (2018) *# Republic.* Princeton university press.

Talbot, C. V., O'Dwyer, S. T., Clare, L., Heaton, J. and Anderson, J. (2020) 'How people with dementia use twitter: A qualitative analysis', *Computers in Human Behavior,* 102, pp. 112-119.

Tarpey, M. (2006) *Why people get involved in health and social care research: a working paper*. Available at: https://www.invo.org.uk/wp-content/uploads/2012/01/whypeoplegetinvolvedinresearch2006.pdf (Accessed: 13 January 2021).

Taylor, C., Gill, L., Gibson, A., Byng, R. and Quinn, C. (2018) 'Engaging “seldom heard” groups in research and intervention development: Offender mental health', *Health Expectations,* 21(6), pp. 1104-1110.

Taylor, M. (2014) 'Information Governance as a Force for Good? Lessons to be Learnt from Care.data', *SCRIPTed,* 11.

Temple, B. (2002) 'Crossed Wires: Interpreters, Translators, and Bilingual Workers in Cross-Language Research', *Qualitative Health Research,* 12(6), pp. 844-854.

Teng, J., Bentley, C., Burgess, M. M., O'Doherty, K. C. and Mcgrail, K. M. (2019) 'Sharing linked data sets for research: results from a deliberative public engagement event in British Columbia, Canada', *International Journal of Population Data Science,* 4(1).

Teodorowski, P. (2020a) 'Micro-level Initiatives to Facilitate the Integration of Resettled Refugees', *Migration Letters,* 17(5), pp. 569-581.

Teodorowski, P. (2020b) 'Migration dissensus among tweeters at #BrexitDay', *Public Policy Studies/Studia z Polityki Publicznej,* 4(28), pp. 83-104.

Teodorowski, P., Ahmed, S. and Kasprzyk, G. (2021) *Making online discussions engaging*. Available at: https://arc-nwc.nihr.ac.uk/uncategorized/making-online-discussions-interactive-and-engaging/ (Accessed: 5 September 2022).

Teodorowski, P., Ahmed, S., Tahir, N., Fleming, K., Rogers, S. and Frith, L. (2022a) 'Involvement and engagement of seldom-heard communities in big data research.', *International Journal of Population Data Science,* 7(3).

Teodorowski, P., Jones, E., Tahir, N., Ahmed, S. and Frith, L. (2021a) 'Public involvement and engagement in big data research: protocol for a scoping review and a systematic review of delivery and effectiveness of strategies for involvement and engagement', *BMJ Open,* 11(8), pp. e050167.

Teodorowski, P., Rodgers, S. E., Fleming, K. and Frith, L. (2022b) 'Use of the Hashtag# DataSavesLives on Twitter: Exploratory and Thematic Analysis', *Journal of medical Internet research,* 24(11), pp. e38232.

Teodorowski, P., Rodgers, S. E., Fleming, K., Tahir, N., Ahmed, S. and Frith, L. (2023a) '‘To me, it's ones and zeros, but in reality that one is death’: A qualitative study exploring researchers' experience of involving and engaging seldom-heard communities in big data research', *Health Expectations,* n/a(n/a).

Teodorowski, P., Rogers, S., Fleming, K., Tahir, N., Ahmed, S. and Frith, L. (2023b) 'Exploring how to improve the involvement of Polish and South Asian communities around big data research. A qualitative study using COM-B model', *International Journal of Population Data Science,* 8(1).

Teodorowski, P., Woods, R., Czarnecka, M. and Kennedy, C. (2021b) 'Brexit, acculturative stress and mental health among EU citizens in Scotland', *Population, Space and Place*.

Thelwall, M. (2013) 'Webometrics and social Web research methods', *University of Wolverhampton,* 140, pp. 140.

Thomas, J. and Harden, A. (2008) 'Methods for the thematic synthesis of qualitative research in systematic reviews', *BMC Medical Research Methodology,* 8(1), pp. 45.

Thompson, J., Barber, R., Ward, P. R., Boote, J. D., Cooper, C. L., Armitage, C. J. and Jones, G. (2009) 'Health researchers' attitudes towards public involvement in health research', *Health Expectations*.

Thornton, H. (1992) 'Breast cancer trials: a patient's viewpoint', *The Lancet,* 339(8784), pp. 44-45.

Thunberg, S. and Arnell, L. (2022) 'Pioneering the use of technologies in qualitative research – A research review of the use of digital interviews', *International Journal of Social Research Methodology,* 25(6), pp. 757-768.

Tierney, S., Dawson, S., Boylan, A.-M., Richards, G., Park, S., Turk, A. and Babatunde, O. (2021) 'Broadening diversity through creative involvement to identify research priorities', *Research Involvement and Engagement,* 7(1).

Tindana, P., de Vries, J., Campbell, M., Littler, K., Seeley, J., Marshall, P., Troyer, J., Ogundipe, M., Alibu, V. P., Yakubu, A. and Parker, M. (2015) 'Community engagement strategies for genomic studies in Africa: a review of the literature', *BMC Med Ethics,* 16, pp. 24.

Todd, S., Coupland, C. and Randall, R. (2020) 'Patient and public involvement facilitators: Could they be the key to the NHS quality improvement agenda?', *Health Expectations,* 23(2), pp. 461-472.

Tomlinson, J., Medlinskiene, K., Cheong, V.-L., Khan, S. and Fylan, B. (2019) 'Patient and public involvement in designing and conducting doctoral research: the whys and the hows', *Research Involvement and Engagement,* 5(1).

Townsend, L. and Wallace, C. (2016) *Social media research: A guide to ethics*. Available at: https://www.gla.ac.uk/media/Media_487729_smxx.pdf (Accessed: 21 February 2020).

Townson, J., Davies, J., Hurt, L., Ashfield-Watt, P. and Paranjothy, S. (2020) 'Developing and evaluating a model of public involvement and engagement embedded in a national longitudinal study: HealthWise Wales', *International Journal of Population Data Science,* 5(3).

Trainor, L. R. and Bundon, A. (2021) 'Developing the craft: reflexive accounts of doing reflexive thematic analysis', *Qualitative Research in Sport, Exercise and Health,* 13(5), pp. 705-726.

Trench, B. (2008) 'Towards an Analytical Framework of Science Communication Models', *Communicating Science in Social Contexts*: Springer Netherlands, pp. 119-135.

Tripathy, J. P., Bhatnagar, A., Shewade, H. D., Kumar, A. M. V., Zachariah, R. and Harries, A. D. (2017) 'Ten tips to improve the visibility and dissemination of research for policy makers and practitioners', *Public Health Action,* 7(1), pp. 10-14.

Tritter, J. Q. (2009) 'Revolution or evolution: the challenges of conceptualizing patient and public involvement in a consumerist world', *Health Expectations,* 12(3), pp. 275-287.

Tritter, J. Q. and Koivusalo, M. (2013) 'Undermining patient and public engagement and limiting its impact:The consequences of the Health and Social Care Act 2012on collective patient and public involvement', *Health Expectations,* 16(2), pp. 115-118.

Troya, M. I., Chew-Graham, C. A., Babatunde, O., Bartlam, B., Higginbottom, A. and Dikomitis, L. (2019) 'Patient and Public Involvement and Engagement in a doctoral research project exploring self-harm in older adults', *Health Expectations,* 22(4), pp. 617-631.

Tully, M. P., Bernsten, C., Aitken, M. and Vass, C. (2020) 'Public preferences regarding data linkage for research: a discrete choice experiment comparing Scotland and Sweden', *BMC Medical Informatics and Decision Making,* 20(1).

Turk, A., McPherson, A., Boylan, A.-M. and Locock, L. (2020) *Researcher's Guide to Patient and Public Involvement*. Available at: https://oxfordbrc.nihr.ac.uk/wp-content/uploads/2017/03/A-Researchers-Guide-to-PPI.pdf (Accessed: 12 January 2021).
Twitter (2022) *Twitter: About*. Available at: https://about.twitter.com/en (Accessed: 15 July 2022).
UK Research Integrity Office (2016) *Good practice in research: Internet-mediated research*.
University of Liverpool 2020. Children Growing Up in Liverpool (C-GULL).
Vaismoradi, M., Turunen, H. and Bondas, T. (2013) 'Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study', *Nursing & Health Sciences,* 15(3), pp. 398-405.
Valdez, R. S., Guterbock, T. M., Thompson, M. J., Reilly, J. D., Menefee, H. K., Bennici, M. S., Williams, I. C. and Rexrode, D. L. (2014) 'Beyond Traditional Advertisements: Leveraging Facebook's Social Structures for Research Recruitment', *Journal of Medical Internet Research,* 16(10), pp. e243.
Vallance, C. (2014) *Adults 'unaware of NHS data plans'*. Available at: https://www.bbc.co.uk/news/health-26187980 (Accessed: 19 February 2023).
Vargas-Silva, C. and Rienzo, C. (2022) *Briefing: Migrants in the UK: An overview*: COMPAS. Available at: https://migrationobservatory.ox.ac.uk/wp-content/uploads/2017/02/MigObs-Briefing-Migrants-in-the-UK-an-overview.pdf.
Vasileiou, K., Barnett, J., Thorpe, S. and Young, T. (2018) 'Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period', *BMC Medical Research Methodology,* 18(1).
Vayena, E. and Blasimme, A. (2017) 'Biomedical Big Data: New Models of Control Over Access, Use and Governance', *Journal of Bioethical Inquiry,* 14(4), pp. 501-513.
Volkmer, A. and Broomfield, K. (2022) *Seldom Heard Voices in Service User Involvement: The how and why of meaningful collaboration.* J & R Press Limited.
Voloder, L. and Kirpitchenko, L. (2016) *Insider research on migration and mobility: international perspectives on researcher positioning.* Routledge.
Waind, E. (2020) 'Trust, security and public interest: Striking the balance', *International Journal of Population Data Science,* 5(3).
Watson, M. and Douglas, F. (2012) 'It's making us look disgusting…and it makes me feel like a mink…it makes me feel depressed!: using photovoice to help 'see' and understand the perspectives of disadvantaged young people about the neighbourhood determinants of their mental well-being', *International Journal of Health Promotion and Education,* 50(6), pp. 278-295.
Weatherall, R. (2019) 'Writing the doctoral thesis differently', *Management Learning,* 50(1), pp. 100-113.
Wegener, C., Meier, N. and Ingerslev, K. (2016) 'Borrowing brainpower – sharing insecurities. Lessons learned from a doctoral peer writing group', *Studies in Higher Education,* 41(6), pp. 1092-1105.
Weich, S., Duncan, C., Bhui, K., Canaway, A., Crepaz-Keay, D., Keown, P., Madan, J., McBride, O., Moon, G., Parsons, H., Singh, S. and Twigg, L. (2018) 'Evaluating the effects of community treatment orders (CTOs) in England using the Mental Health Services Dataset (MHSDS): protocol for a national, population-based study', *BMJ Open,* 8(10).
Welsh, E. 'Dealing with data: Using NVivo in the qualitative data analysis process'. *Forum qualitative sozialforschung/Forum: qualitative social research*.
West, R., Michie, S., Rubin, G. J. and Amlôt, R. (2020) 'Applying principles of behaviour change to reduce SARS-CoV-2 transmission', *Nature Human Behaviour,* 4(5), pp. 451-459.
Whitaker, C., Stevelink, S. and Fear, N. (2017) 'The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review', *Journal of medical Internet research,* 19(8), pp. e290-e290.
Wiggins, S. (2016) 'Discursive psychology: Theory, method and applications', *Discursive Psychology*, pp. 1-280.

Wilkinson, S. (1988) 'The role of reflexivity in feminist psychology', *Women's Studies International Forum,* 11(5), pp. 493-502.

Williams, M. L., Burnap, P. and Sloan, L. (2017) 'Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation', *Sociology,* 51(6), pp. 1149-1168.

Willison, D. J., Trowbridge, J., Greiver, M., Keshavjee, K., Mumford, D. and Sullivan, F. (2019) 'Participatory governance over research in an academic research network: The case of Diabetes Action Canada', *BMJ Open,* 9(4).

Willmott, T. J., Pang, B. and Rundle-Thiele, S. (2021) 'Capability, opportunity, and motivation: an across contexts empirical examination of the COM-B model', *BMC Public Health,* 21(1).

Wise, J. (2022) 'New data strategy for England aims to rebuild public trust', *BMJ,* 377, pp. o1455.

Woolfall, K. (2023) 'Identifying and preventing fraudulent participation in qualitative research', *Archives of Disease in Childhood,* 108(6), pp. 421-422.

Wren, Y. and CLAPA Cleft Collective Patient Consultation Group (2022) 'Working with people across the lifespan', in Volkmer, A. and Broomfield, K. (eds.) *Seldom Heard Voices in Service User Involvement: The how and why of meaningful collaboration*: J & R Press Limited, pp. 61-70.

Wynne, B. (2006) 'Public Engagement as a Means of Restoring Public Trust in Science – Hitting the Notes, but Missing the Music?', *Community genetics,* 9, pp. 211-20.

Xafis, V. and Labude, M. K. (2019) 'Openness in Big Data and Data Repositories', *Asian Bioethics Review,* 11(3), pp. 255-273.

Zhang, X., Pérez-Stable, E. J., Bourne, P. E., Peprah, E., Duru, O. K., Breen, N., Berrigan, D., Wood, F., Jackson, J. S., Wong, D. W. S. and Denny, J. (2017) 'Big Data Science: Opportunities and Challenges to Address Minority Health and Health Disparities in the 21st Century', *Ethnicity & Disease,* 27(2), pp. 95.

Zhu, X., Kim, Y. and Park, H. (2020) 'Do messages spread widely also diffuse fast? Examining the effects of message characteristics on information diffusion', *Computers in Human Behavior,* 103, pp. 37-47.

Zoccatelli, G., Desai, A., Martin, G., Brearley, S., Murrels, T. and Robert, G. (2020) 'Enabling 'citizen voice' in the English health and social care system: A national survey of the organizational structures, relationships and impacts of local Healthwatch in England', *Health Expectations,* 23(5), pp. 1108-1117.

Zoom (2021) *Security Guide*. Available at: https://zoom.us/docs/doc/Zoom-Security-White-Paper.pdf (Accessed: 20 August 2021).

Appendices

Appendix 1: Data extraction form

Title:

Year:

Author:

Number:

Generic Information

Paper aim	
Design (e.g. qualitative, discussion paper)	
Country (if specified)	
Context (e.g. health condition)	
Demographics of participants for PPI (also record if there are a seldom-heard group based if paper defines them as such).	
Methods (if relevant)	
Funding	

PPIE

Is It involvement, engagement or consultation (based on NIHR definition)?	
Process of involvement or engagement (e.g. advisory board, co-researchers etc.)	
Legal or ethical issues (in relation to PPI)	
References to PPI guidance and policies	
Challenges of PPI	
Facilitators of PPI	
Key Message	
Other relevant information	

Does the paper discuss the delivery or effectiveness of involvement and engagement strategies?

Yes **No**

Include in systematic review?

Yes **No**

Appendix 2: Topic guide for interviews with Polish & South Asians

Interview schedule Polish & South Asians v.2 26.8.21

Main questions are numbered with optional prompts in italics.

Beginning

- Welcome participant, and thank them for interest
- Make sure they received the participant information sheet and signed previously sent consent form, invite questions. Remind that the session will last around 60 minutes.
- Start recording the meeting. Remind about recording and consent.
- Remind participants:
    - They don't have to answer any question that they do not wish to respond to, and they can stop the interview without giving any reason.
    - There are no right or wrong answers as we hope to understand their views and perceptions.
    - It is okay to have a break during the interview if they wish to.

Warm-up discussion (providing background to interviewees involvement in the community)

1. How long how have you been living in the area?
   How long have you lived in the UK? (if migrant)

2. Could you tell me if and how are you involved in your community?
   Are you a member of any local charity or groups?

Data collection & usage

Provide description to what big data is.

Let me explain what does the big data mean. Over the last two decades, the ongoing digitalisation of information has allowed the creation and linkage of large, multi-source health data sets to provide novel healthcare applications (e.g. research of new medicine, evaluation of service and statistical purposes). This is often called 'big data'. The term refers to datasets that are too large or complex to process using traditional methods. They can be created by healthcare professionals or patients. The data can be structured (e.g. statistical data) or unstructured (e.g. observations, medical notes). These include medical health records, health-related databases, patient monitoring their health through phone applications. Researchers, policy-makers or private companies can use big data. Patient data can be anonymised or pseudonymised (an individual can still be identified through additional information).

Further examples of big data projects can be found at Data Saves Lives campaign website.

3. How you ever been interested in how your medical data is being (re)used?
   How do you fell about how it is used?

4. What kind of medical data do you feel researchers should have access?
   How long should the data be accessible?
   Where should it be stored?

5. Who should have access to your medical data?
   Medical professionals, researchers, private companies, government, transfer abroad?

Public involvement

6. How do you think researchers should involve communities in the design and governance of research?
   Anything particular within your Polish/South Asian communities?
   Where should researchers recruit people?
   Who should get involved? How to make sure they are involved?

7. What are the challenges or things helping you to involve your community in the design and governance of research?
   What are their causes?
   How can these barriers be overcome? Who is responsible for it?

Public engagement

8. How much are you interested in learning about how your medical data is used for research?
   Would you like to know what changes and new solutions for medicine these offers?
   Is it important to do it?

9. How do you feel researchers should share and promote research projects?
   Anything particular within your Polish/South Asian communities?
   What is the best medium to share this information? Tv, newspapers, NHS leaflets or social media?
   What about the language barrier?
   Have you been exposed to any campaigns around benefits of big data research? E.g. #DataSavesLives

Conclusions

10. What do you feel should happen because of this research?
    How could it be shared among researchers and seldom-heard communities?

11. These were all questions from me. Would you like to add anything else to what we discussed?

Wrap-up

- Stop the recording
- Thanks to the interviewee.
- Ask if they would like to receive a copy of the research findings.
- Information about a voucher and say if they have any problem with it, they can contact the researcher directly to look into it.
- If based in ARC NWC area, provide information about ARC Public Advisor Network.

Appendix 3: Topic guide for interviews with researchers
Interview schedule researchers v.3 26.01.22

Main questions are numbered with optional prompts in italics.

Beginning

- Welcome participant and thank them for interest.
- Make sure they received the participant information sheet and signed previously sent consent form, invite questions. Remind that the session will last around 30 minutes.
- Start recording the meeting. Remind about recording and consent.
- Remind participants:
    - They don't have to answer any question that they do not wish to respond to, and they can stop the interview without giving any reason.
    - There are no right or wrong answers as we hope to understand their views and perceptions.
    - It is okay to have a break during the interview if they wish to.

Warm-up discussion

Say what we mean in this study as public involvement & engagement – following Involve definition. Definitions remain disputed so providing these definitions will assist in settling the scene.

<u>Public involvement</u> – "*research being carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them.*"

<u>Public engagement</u> – "*information and knowledge about research is provided and disseminated.*" – this usually takes place after the project is concluded.

1. Could you tell me more about the research you conduct/conducted or consider that involve or engage with seldom heard communities?
   How would you define a community as seldom heard??

2. Which seldom heard communities did you involved or engaged /consider involving or engaging with (in the past)?
   How did you identify these communities as seldom heard groups?
   What was your relationship with these communities?
   Are there any groups which you haven't engaged but you would like to? Why not?

Experience of public involvement & engagement

3. How easy will be/was/is to reach seldom-heard communities in your research?

4. Are there barriers to involving & engaging seldom heard communities in big data research?
   What causes these barriers?
   How can these barriers be limited?
   How easy was to explain your project in plain language?

5. Are there enablers in involving & engaging seldom heard communities in big data research?
   Did you work with any local gatekeepers?
   How did you recruit public members involved in your study?
   Have you provided any training for public members?
   How easy was to make the involvement user-friendly?
   Have you utlisied already ongoing campaigns such as #DataSavesLives?

6. What is the impact of involvement and engagement with seldom heard communities?
   Was it different from the impact of engagement with seldom heard communities on the study?
   Do you feel that this made a change? If yes, how?

Conclusions

7. What do you feel should happen because of our research?
   How could it be shared among researchers and seldom-heard communities?
   How could researchers be supported better to reach seldom heard communities?

8. These were all questions from me. Would you like to add anything else to what we discussed?

Wrap-up

- Stop the recording.
- Thanks to the interviewee.
- Ask if they would like to receive a copy of the research findings.

Appendix 4: Example of participant information sheet

Participant information sheet: interviews- South Asian & Polish

Version number & date: v.2 26/08/2021

Research ethics application number: 10063

Title of the research project: Exploring public involvement and engagement of seldom heard communities in big data research.

Invitation

You are being invited to participate in this research project. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask us if you would like more information or if there is anything that you do not understand. Please also feel free to discuss this with your colleagues, friends or relatives if you wish. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to.

Thank you for reading this.

1. Why are we doing this study?

This project is part of a PhD project titled: Public involvement and engagement of seldom heard communities in big data research. The aim of the project is to explore how to involve seldom heard communities in public involvement and engagement around big data research.

2. Why have I been asked to take part?

You have been asked to take part because we would like to ask for your views on views and perceptions on health data and how public members can get involved and engaged in research.

3. Do I have to take part?

No. You do not have to take part, participating is entirely voluntary.

4. What will happen if I take part?

An interview – this will be done online or over the telephone and the audio will be recorded with your permission. The interviews will take approximately 60 minutes. If you would like to, we can arrange an interpreter in the language of your choice to support the discussion.

5. How will my data be used?

The University processes personal data as part of its research and teaching activities in accordance with the lawful basis of ‘public task’, and in accordance with the University’s purpose of “advancing education, learning and research for the public benefit.

Under UK data protection legislation, the University acts as the Data Controller for personal data collected as part of the University’s research. The Principal Investigator acts as the Data Processor for this study, and any queries relating to the handling of your personal data can be sent to Dr Lucy Frith.

Further information on how your data will be used can be found in the table below.

How will my data be collected?	Copy of the initialled consent form stored separately from audio recordings and transcribed interviews. Initially an audio recording of the interview will be stored. It will then be transcribed and anonymised. The original recording will then be deleted. The transcribed anonymized interview will be stored.
How will my data be stored?	All data will be stored on the password protected University server.
How long will my data be stored for?	The audio will be deleted after it has been transcribed. The transcribed anonymized interview will be kept for 10 years
What measures are in place to protect the security and confidentiality of my data?	All research results will be fully anonymised, all data shared in the research team will be anonymised.
Will my data be anonymised?	Yes
How will my data be used?	The interview data will form a part of a PhD thesis. I may write articles or give talks about the research findings. Any quotes from the interviews will be anonymized.
Who will have access to my data?	The research team will have access to the anonymized transcribed data. The sample of interpretered or translated interviews will be checked by the second interpreter or translator to assess the accuracy of interpreting or translation. Data will be prepared and fully anonymised data deposited for sharing on the UK Data Service (UKDS) ReShare.
Will my data be archived for use in other research projects in the future?	Yes
How will my data be destroyed?	Deleted from the University server.

6. Are there any risks in taking part?

We do not think that there are any risks to taking part. Although the interview will take up some of your time and we will ask about your views. We do not intend to cover your experiences or views outside of the remit of this study. If any of the issues discussed are distressing, or you don’t want to answer them, you are free to stop the interview. You do not have to answer any questions that you feel uncomfortable with and you don’t have to explain your decision.

7. Are there any benefits in taking part?

There are no direct benefits to you as an individual participating in this project, but this project will hopefully help us to better understand how involve and engage with seldom heard communities in big data research.

8. Will I be reimbursed for participating in this study?

Yes, upon completion of the interview, you will receive a £20 for Amazon or high street shopping voucher.

9. What will happen to the results of the study?

The interview data will form part of a PhD thesis. The thesis will be open access available on the University of Liverpool website. The findings might be included in academic journal articles and they might be presented at events and conferences.

Please contact me (p.teodorowski@liverpool.ac.uk) if you are interested in receiving updates regarding the progress and findings of the study.

10. What will happen if I want to stop taking part?

You can discontinue the interview at any point without giving a reason. The data can be withdrawn from the study up to two weeks after the interview, then the data will be transcribed and anonymised and you can no longer withdraw your data.

11. What if I am unhappy or if there is a problem?

If you are unhappy, or if there is a problem, please feel free to let us know by contacting Piotr Teodorowski p.teodorowski@liverpool.ac.uk or Lucy Frith, frith@liverpool.ac.uk 0151-795-5333, and we will try to help.

If you remain unhappy or have a complaint which you feel you cannot come to us with then you should contact the Research Ethics and Integrity Office at ethics@liv.ac.uk. When contacting the Research Ethics and Integrity Office, please provide details of the name or description of the study (so that it can be identified), and the researcher(s) involved, and the details of the complaint you wish to make.

The University strives to maintain the highest standards of rigour in the processing of your data. However, if you have any concerns about the way in which the University processes your personal data, it is important that you are aware of your right to lodge a complaint with the Information Commissioner's Office by calling 0303 123 1113.

12. Who can I contact if I have further questions?

p.teodorowski@liverpool.ac.uk

Appendix 5: Approval letter from ethics committee for Twitter study

UNIVERSITY OF
LIVERPOOL

Institute of Population
Health Research Ethics
Committee

27 April 2021

Dear Dr Frith

I am pleased to inform you that your application for research ethics approval has been approved. Application details and conditions of approval can be found below. Appendix A contains a list of documents approved by the Committee.

Application Details

Reference: 9815
Project Title: Exploring the public engagement campaign #DataSavesLives on Twitter.
Principal Investigator/Supervisor: Dr Lucy Frith
Co-Investigator(s): Mr Piotr Teodorowski
Lead Student Investigator: -
Department: Public Health, Policy & Systems
Approval Date: 27/04/2021
Approval Expiry Date: Five years from the approval date listed above

The application was **APPROVED** subject to the following conditions:

Conditions of approval

Please note: this approval is subject to the University's research restrictions during the pandemic, as laid out on the research ethics webpages. Therefore, wherever possible, research should be conducted via remote means which avoid the need for face-to-face contact with human participants during the pandemic. The process for requesting an exemption to these restrictions is described on the research ethics webpages.

- All serious adverse events must be reported to the Committee (ethics@liverpool.ac.uk) in accordance with the procedure for reporting adverse events.
- If you wish to extend the duration of the study beyond the research ethics approval expiry

date listed above, a new application should be submitted.

- If you wish to make an amendment to the study, please create and submit an amendment form using the research ethics system. If the named Principal Investigator or Supervisor changes, or leaves the employment of the University during the course of this approval, the approval will lapse. Therefore it will be necessary to create and submit an amendment form within the research ethics system.
- It is the responsibility of the Principal Investigator/Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Institute of Population Health Research Ethics Committee iphethics@liverpool.ac.uk

IPH-REC

Appendix - Approved Documents

(Relevant only to amendments involving changes to the study documentation) The final document set reviewed and approved by the committee is listed below:

Appendix 6: Approval letter from ethics committee for interviews with seldom-heard communities and researchers

Institute of Population
Health Research Ethics
Committee

8 September 2021

Dear Dr Frith

I am pleased to inform you that your application for research ethics approval has been approved. Application details and conditions of approval can be found below. Appendix A contains a list of documents approved by the Committee.

Application Details

Reference: 10063
Project Title: Exploring public involvement and engagement of
seldom heard communities in big data research Principal
Investigator/Supervisor: Dr Lucy Frith
Co-Investigator(s): Mr Piotr Teodorowski,
Ms Sarah Rogers, Dr Kate Fleming Lead Student
Investigator: -
Department: Public Health, Policy & Systems
Approval Date: 08/09/2021
Approval Expiry Date: Five years from the approval date listed above

The application was **APPROVED** subject to the following conditions:

Conditions of approval

Please note: this approval is subject to the University's research restrictions during the pandemic, as laid out on the research ethics webpages. Therefore, wherever possible, research should be conducted via remote means which avoid the need for face-to-face contact with human participants during the pandemic. The process for requesting an exemption to these restrictions is described on the research ethics webpages.

- All serious adverse events must be reported to the Committee (ethics@liverpool.ac.uk) in accordance with the procedure for reporting adverse events.

- If you wish to extend the duration of the study beyond the research ethics approval expiry date listed above, a new application should be submitted.
- If you wish to make an amendment to the study, please create and submit an
- amendment form using the research ethics system. If the named Principal Investigator or Supervisor changes, or leaves the employment of the University during the course of this approval, the approval will lapse. Therefore it will be necessary to create and submit an amendment form within the research ethics system.
- It is the responsibility of the Principal Investigator/Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Institute of Population Health Research Ethics

Committee iphethics@liverpool.ac.uk

IPH-REC

Appendix - Approved Documents

(Relevant only to amendments involving

changes to the study documentation) The

final document set reviewed and approved by

the committee is listed below:

Document Type	File Name	Date	Version
Participant Consent Form	Consent form Interviews Polish South Asian v1 06.5.21 - Copy	06/05/2021	1
Participant Consent Form	Consent form Interviews Researchers v1 06.5.21	06/05/2021	1
Advertisement	Ad researchers	02/08/2021	1
Advertisement	Ad Polish South Asians v2	26/08/2021	2
Interview Schedule	Interview schedule Polish South Asians	26/08/2021	2
Interview Schedule	Interview schedule researchers	26/08/2021	2
Participant	Information sheet interviews Researchers	26/08/	2

Information Sheet		2021	
Participant Information Sheet	Information sheet interviews Polish South Asian	26/08/ 2021	2

Appendix 7: Public contributors call shared with the Public Advisors Forum

Invitation for an opportunity for public advisors in the ARC

Subject: Expression of Interest for public involvement in PhD study

Dear Adviser,

I am an ARC funded PhD student at the University of Liverpool, exploring public involvement of seldom heard voices in managing and organising large health data sets and would like ongoing public involvement throughout my research journey. The aim of the involving public advisors is to get your views on various stages of my research including (but not limited to) review protocol, initial review findings, study design and participants recruitment.

I am seeking up two public advisors to participate in the first event focusing on the design of the review protocol. No prior skills in conducting systematic reviews are needed. The discussion will focus on review aims and defining what could be seldom heard voices in the context of big data research. Here you can find a short introduction to big data in healthcare. The meeting will take place in late November, and time will be arranged with public advisors interested in contributing.

Due to Covid-19 situation, the workshop will be held online through Zoom and take up to 2 hours. There will be a £25 reimbursement for contributions at the workshop. No further meetings are planned this year, and there will be 3-4 meetings the next year. Time will be arranged directly with involved public advisors to fit your schedule – I am open to evenings or weekend if this will be the most suitable time for you.

If you are interested, please reply with a short paragraph about why you would like to participate. In the event of more than two advisers expressing an interest, first preference will go to those who consider themselves members of seldom heard groups in big data research and would like to contribute throughout all three years of my PhD.

If you would have any questions or would like to discuss this opportunity before applying, please do not hesitate to email me.

Expressions of interest to be received no later than Monday 16th November and sent to p.teodorowski@liverpool.ac.uk

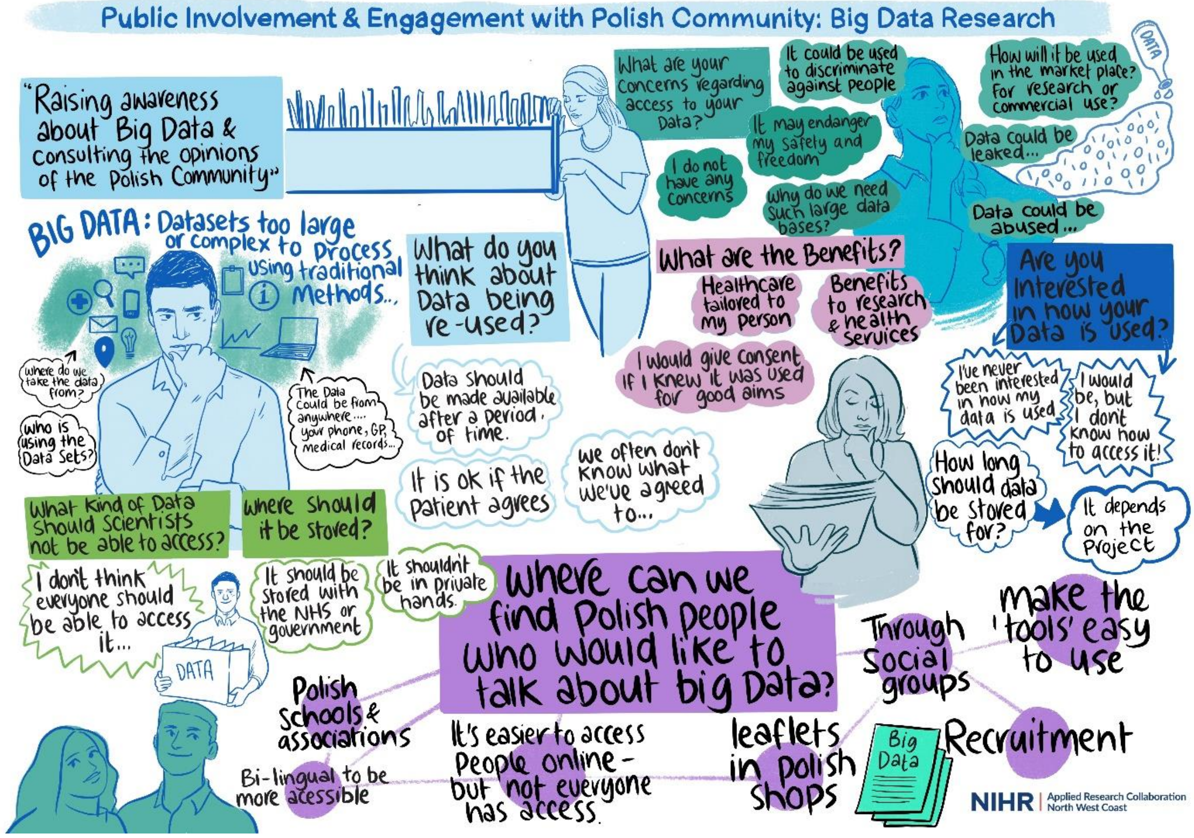
Best Wishes,

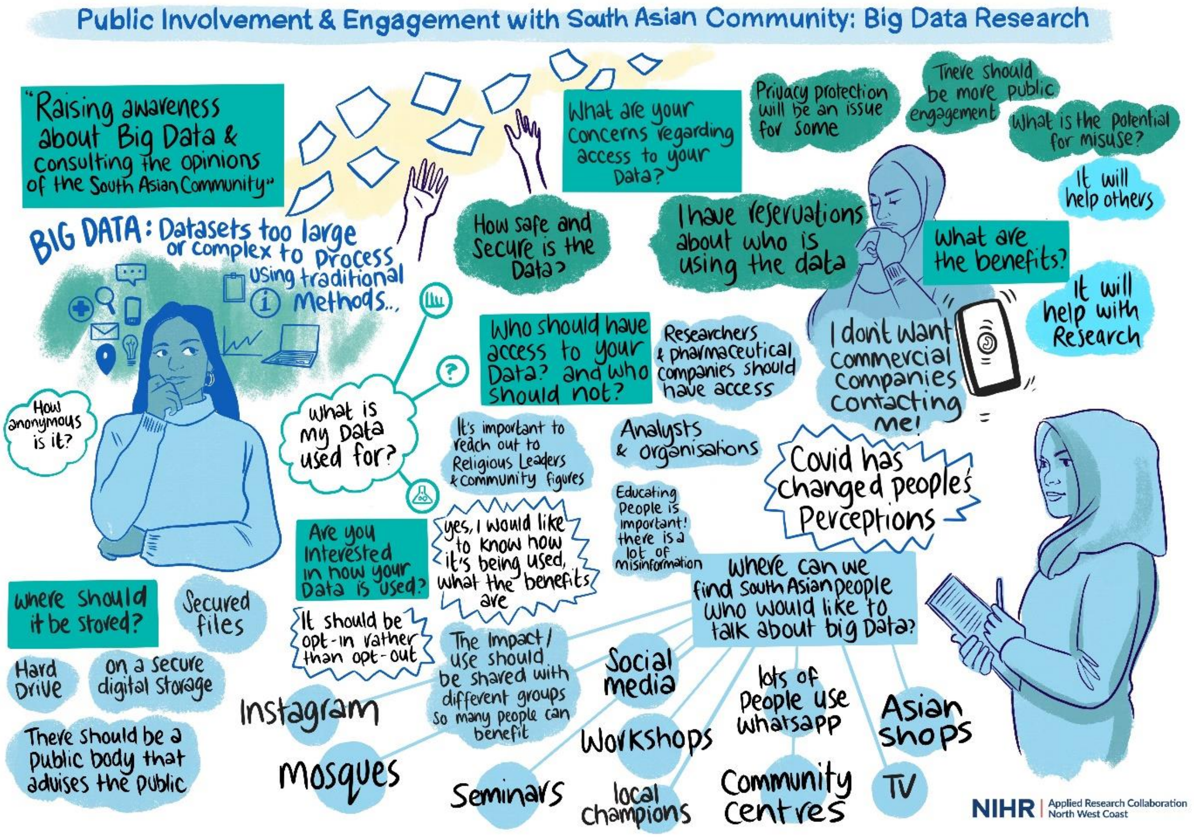
Piotr Teodorowski

Rough fee structure

2-hour meeting plus preparation (i.e. management board of a theme	£50
2-hour meeting attendance (no preparation)	£25

Appendix 8: Visual minutes from public involvement sessions





Appendix 9: Further examples of involvement and engagement in big data research from researcher's perspectives.

Supporting quotes of how seldom-heard communities were engaged in big data research according to theme

Theme 1

"There's lots of things that people might not have heard of before, even considered or understood that goes on behind the scenes and some very complex areas to discuss, and we've had a few things that have been quite difficult (…) like trusted research environments (…) So it's definitely taken time to build up the knowledge base of [public contributors], but we've had, for example, external speakers from other organisations that work with this quite often (…) we share training or webinars with [public contributors] that might be of interest and if we're going to have a discussion on a particular area, we'll tend to kind of give them some background reading or a paper from us or a video to go and watch or something to help build their knowledge and we also tried to have discussions over a couple of meetings that we don't just have one discussion about it and that's it. We try to kind of embed it in the process a bit earlier so that we can have a few discussions, so again gives them more of a chance to kind of get grips with it and understand a bit better." (Arabella)

"This was at the request of the patient and public contributors on that particular project, they said they really wanted to get to grips with some of the statistical methods that we were using (…) they really felt that if they understood the methods, they might be able to have a more fruitful conversation with us about how they feel about the way that we're running the study and how we're designing the analysis and things like that. So we ran a session where the one of the analysts on the project and I worked with them on the presentation. They gave a presentation to the patient and public contributors which walked through everything from what the GP will put into their computer and when they're having a consultation with them right through to where that data goes, what it looks like. Obviously, it was all just anonymized examples. It wasn't actual, they couldn't see the actual data. Then how the different methods that they use to process it and how that method is applied in the in the specific project and the patient and public contributors said, 'actually, the fact that you've taken the time to sit down and explain that to us and say specifically how it's going to be used in this project and was really useful' and we had a actually a really interesting discussion about could you include this or do you exclude this group of patients (…) and actually they started to come up with things which were coming from their own experience because they understood enough from what they said they could really add to the conversation, so I think taking the time and making the effort, if the contributors asked for it was really helpful." (Sienna)

Theme 2

"*We went as wide as we could because we really did want geographical location as well as diversity (…) our key things were lived experience of health inequalities. Then we looked as to access to maybe wider networks and then making sure that we had ethnic diversity across the range. We were able to recruit people from African backgrounds, Asian backgrounds and then we have some members who are white but have a different religion so Jewish for example and also [health conditions] (…) we've got quite a range, a diverse range of people, but we do recognise that the one individual cannot represent the whole population*"(Penelope)

"If you want to hear from these groups, then you need to reward them and you should reward anyway, but you know, obviously resources are slim (…) it's about picking your targets, making sure that what you're doing suits the study that you're looking for. Because you know we don't

have infinite resources, so you have to basically use them wisely. And if it's a study that is important to seldom heard groups, then you know you would need to target that and put your resources into that, and that's like finding out where they are, what they do, what things are important to them and going out to them there." (August)

Theme 3

"There are links to other organisations, for example, [local charity working with seldom-heard community], and so we've had a really positive conversation with them and there could be mutually beneficial opportunities for us to open the invitation to people that they support to be part of our [public involvement group] and in turn it kind of supports their personal development programmes that they do for people (…) so we get the benefit of expanding the perspectives that are included within our group, but hopefully that's a useful thing for their personal development as well as part of the programmes that are doing underneath the charity umbrella." (Alex)

Theme 4

"I think to facilitate good discussion. I think that you want no more than like 10 people. Because otherwise you can't include everybody and people become observers to a more dominant group, so (…) I think what works well for and even face-to-face, we would do like a smaller group so to ensure that everybody can contribute." (August)

"They become experts, but you can't help that if you're in that realm. We have ways of mitigating that, and might be that naturally members leave, then you recruit newer members. But then there's so many challenges around that it's training. There has to be a level of training and understanding and does that make just because somebody got some knowledge. Does that make them make their opinion less worthy? I think, as long as we sense check it to the wider population and go out to community groups. (Penelope)

Appendix 10 Standards for Reporting Qualitative Research (SRQR)*

	Standards for Reporting Qualitative Research (SRQR)*	
	http://www.equator-network.org/reporting-guidelines/srqr/	
		Page/line no(s).
Title and abstract		
	Title - Concise description of the nature and topic of the study Identifying the study as qualitative or indicating the approach (e.g., ethnography, grounded theory) or data collection methods (e.g., interview, focus group) is recommended	Page 192
	Abstract - Summary of key elements of the study using the abstract format of the intended publication; typically includes background, purpose, methods, results, and conclusions	Page 193
Introduction		
	Problem formulation - Description and significance of the problem/phenomenon studied; review of relevant theory and empirical work; problem statement	Page 194
	Purpose or research question - Purpose of the study and specific objectives or questions	Page 197
Methods		
	Qualitative approach and research paradigm - Qualitative approach (e.g., ethnography, grounded theory, case study, phenomenology, narrative research) and guiding theory if appropriate; identifying the research paradigm (e.g., postpositivist, constructivist/ interpretivist) is also recommended; rationale**	Page 197
	Researcher characteristics and reflexivity - Researchers' characteristics that may influence the research, including personal attributes, qualifications/experience, relationship with participants, assumptions, and/or presuppositions; potential or actual interaction between researchers' characteristics and the research questions, approach, methods, results, and/or transferability	Page 200
	Context - Setting/site and salient contextual factors; rationale**	Page 197
	Sampling strategy - How and why research participants, documents, or events were selected; criteria for deciding when no further sampling was necessary (e.g., sampling saturation); rationale**	Page 197
	Ethical issues pertaining to human subjects - Documentation of approval by an appropriate ethics review board and participant consent, or explanation for lack thereof; other confidentiality and data security issues	Page 200
	Data collection methods - Types of data collected; details of data collection procedures including (as appropriate) start and stop dates of data collection and analysis, iterative process, triangulation of sources/methods, and modification of procedures in response to evolving study findings; rationale**	Page 198

	Data collection instruments and technologies - Description of instruments (e.g., interview guides, questionnaires) and devices (e.g., audio recorders) used for data collection; if/how the instrument(s) changed over the course of the study	Page 198
	Units of study - Number and relevant characteristics of participants, documents, or events included in the study; level of participation (could be reported in results)	Page 197
	Data processing - Methods for processing data prior to and during analysis, including transcription, data entry, data management and security, verification of data integrity, data coding, and anonymization/de-identification of excerpts	Page 200
	Data analysis - Process by which inferences, themes, etc., were identified and developed, including the researchers involved in data analysis; usually references a specific paradigm or approach; rationale**	Page 200
	Techniques to enhance trustworthiness - Techniques to enhance trustworthiness and credibility of data analysis (e.g., member checking, audit trail, triangulation); rationale**	Page 200
Results/findings		
	Synthesis and interpretation - Main findings (e.g., interpretations, inferences, and themes); might include development of a theory or model, or integration with prior research or theory	Page 200
	Links to empirical data - Evidence (e.g., quotes, field notes, text excerpts, photographs) to substantiate analytic findings	Page 200
Discussion		
	Integration with prior work, implications, transferability, and contribution(s) to the field - Short summary of main findings; explanation of how findings and conclusions connect to, support, elaborate on, or challenge conclusions of earlier scholarship; discussion of scope of application/generalizability; identification of unique contribution(s) to scholarship in a discipline or field	Page 211
	Limitations - Trustworthiness and limitations of findings	Page 213
Other		
	Conflicts of interest - Potential sources of influence or perceived influence on study conduct and conclusions; how these were managed	
	Funding - Sources of funding and other support; role of funders in data collection, interpretation, and reporting	
	*The authors created the SRQR by searching the literature to identify guidelines, reporting standards, and critical appraisal criteria for qualitative research; reviewing the reference lists of retrieved sources; and contacting experts to gain feedback. The SRQR aims to improve the transparency of all aspects of qualitative research by providing clear standards for reporting qualitative research.	
	**The rationale should briefly discuss the justification for choosing that theory, approach, method, or technique rather than other options available, the assumptions and limitations implicit in those choices, and how those choices influence study conclusions and transferability. As appropriate, the rationale for several items might be discussed together.	
	Reference:	

	O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine*, Vol. 89, No. 9 / Sept 2014 DOI: 10.1097/ACM.0000000000000388	