

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

Exploring how to improve the involvement of Polish and South Asian communities around big data research. A qualitative study using COM-B model

Piotr Teodorowski^{1,*}, Sarah E. Rodgers¹, Kate Fleming², Naheed Tahir³, Saiqa Ahmed³, and Lucy Frith⁴

Submission History

Submitted:	30/11/2022
Accepted:	23/05/2023
Published:	13/07/2023

¹Department of Public Health, Policy & Systems, University of Liverpool, Liverpool, UK, L69 3GB

²National Disease Registration Service, NHS England, Leeds, West Yorkshire, LS1 4AP

³ARC NWC Public Advisor, Liverpool, UK, L69 3GB

⁴Centre for Social Ethics and Policy, University of Manchester, Manchester, UK, M13 9QQ

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 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.

Keywords

PPI; public involvement; qualitative; big data; ethnic minorities

*Corresponding Author:

Email Address: p.teodorowski@liverpool.ac.uk (Piotr Teodorowski)



Introduction

Involving the public in research design and execution has become firmly embedded within the UK academic research environment [1]. Public involvement has had a positive impact on research [2] and moral and ethical principles support public involvement [3]. It offers the researcher a lay perspective, especially from patients who have a lived experience of the health conditions being studied [4] and has the potential to improve the quality of research. In this paper we focus on public involvement in big data research [5]. There remain disagreements about what constitutes big data (also called data-intensive research [6] or routinely collected data [7]); 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” [8]. 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 [9]. The public might feel disconnected from research [5] 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 [10]. Ensuring social licence for research, which is more than meeting legal requirements requires trustworthiness and addressing public concerns is important [11, 12]. 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 [13–15]. Public contributors can also take part in public deliberations on linking data for research or its management [16, 17], co-design ways of engaging with the broader public [18] and be a public representative (contributor) on the project's advisory group [7].

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 [8]. 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 experience health inequalities [19]. Researchers and policymakers must include diverse voices to ensure that services can effectively respond to all communities' needs [20]. 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 [21]) Public involvement can help in formulating initiatives to reach seldom-heard communities [22].

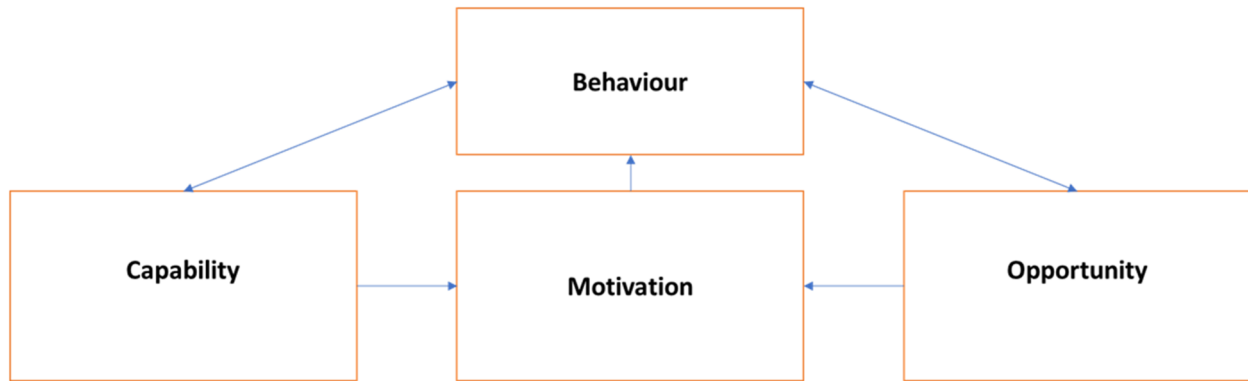
Despite recognising the importance of these kinds of initiatives, there is limited literature on how to successfully involve seldom-heard communities in health and social care research [23, 24], and there is a particular lack of literature on how to involve people in big data research. A bibliometric review by Boote and colleagues [25] 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 [25]. 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 [17, 18], and training on technical terms might be required [26].

People's attitudes towards being involved in research could be influenced by personal and social motivations [27, 28], including their available time. Additionally, involvement requires additional time and resources from the researcher [29]. 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.

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. 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 in the former the responsibility is on researchers [21]. 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 [30]. 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. However, it does not mean that researchers would need to equally influence each of them, even if they can.

Figure 1: COM-B model as developed by Michie and colleagues [30]



We adopted the COM-B model (Figure 1) 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 [31]. 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 [32, 33] and to understand health behaviour change [34–36].

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.

Methods

We used the Standards for Reporting Qualitative Research [37] to ensure transparency when reporting this research. The completed guidance is available in Appendix 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. They used to be the largest migrant community (based on the country of birth) in the UK until 2018 [38]. 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 [39] and Brexit as a political event negatively influenced Poles' mental health and wellbeing [40].

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 [41–43]. 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 [44].

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.

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 ads. 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 [45]. 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 [46]. We followed Gawlewicz's [44] 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.

Public contributors

Involving public contributors in the qualitative analysis can improve the quality of findings [47]. 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 public 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 [48]. For example, they were able to point out nuances in the initial interview that led to additional questions for the following interviews.

Data analysis

All interviews were uploaded to NVivo 12 and analysed using reflexive thematic analysis [49–51]. 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 [52]. 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.

Ethics

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

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 2). The themes appeared across both participant groups, indicating the results could be applicable to other migrant and ethnic minority groups in the UK.

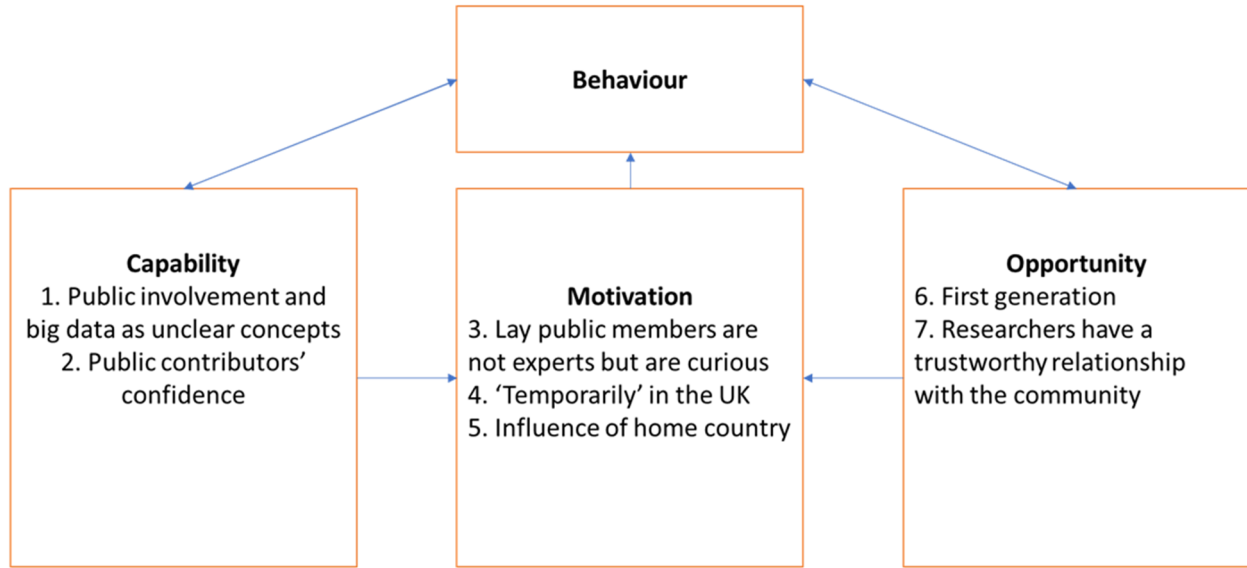
Happy to reuse data

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

Figure 2: Themes mapped under COM-B model



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 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 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 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 should be allowed to withdraw at any time. You know that 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.

Capability

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 it 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.

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 public 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.

Motivation

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, , 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 elderly 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.

'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.

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.

Opportunity

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.

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)

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 involved 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 [53]. 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. This 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 may be challenging and time-consuming for researchers [54]. 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) [55].

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 [56, 57]. 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 [58].

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 [11]. 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 [59]. 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 [60]. 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 [61].

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 [39]) 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 [40].

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 [16], 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 [63] 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 [58]. Even among researchers who support public involvement, not all consider every research stage appropriate for public involvement [47]. Although big data research can be a complex topic, research can (and should) make it inclusive and accessible to members of the public [5].

Education and training are needed for successful public involvement [9], 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 [13]. Involving public members with technical skills can be a solution, but the danger is that, as Kirkham and colleagues [63] 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.

Strengths and limitations

Our study is the first to explore how to improve the involvement 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.

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.

Acknowledgements

This report is independent research funded by the National Institute for Health Research Applied Research Collaboration North West Coast (ARC NWC). The views expressed in this publication are those of the author(s) and not necessarily

those of the National Institute for Health Research or the Department of Health and Social Care. SER is partly funded by the National Institute for Health Research (NIHR) Applied Research Collaboration North West Coast.

Statement on conflicts of Interest

None declared

Ethics Statement

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

References

1. NIHR. National Standards for Public Involvement. 2019.
2. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. *Health Expect*. 2014;17(5):637-50. Epub 2012/07/19. <http://doi.org/10.1111/j.1369-7625.2012.00795.x>. PubMed PMID: 22809132.
3. Thompson J, Barber R, Ward PR, Boote JD, Cooper CL, Armitage CJ, et al. Health researchers' attitudes towards public involvement in health research. *Health Expectations*. 2009. <http://doi.org/10.1111/j.1369-7625.2009.00532.x>
4. Rhodes P, Nocon A, Booth M, Chowdrey MY, Fabian A, Lambert N, et al. A service users' research advisory group from the perspectives of both service users and researchers. *Health & Social Care in the Community*. 2002;10(5):402-9. <https://doi.org/10.1046/j.1365-2524.2002.00376.x>
5. Aitken M, Tully M, Porteous C, Denegri S, Cunningham-Burley S, Banner N, et al. Consensus Statement on Public Involvement and Engagement with Data-Intensive Health Research. *International Journal of Population Data Science*. 2019;4. <http://doi.org/10.23889/ijpds.v4i1.586>
6. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics*. 2018;114:57-65. <http://doi.org/10.1016/j.ijmedinf.2018.03.013>
7. Newburn M, Scanlon M, Plachcinski R, Macfarlane AJ. 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*. 2020;5(3). <http://doi.org/10.23889/ijpds.v5i3.1366>
8. NIHR. Briefing notes for researchers - public involvement in NHS, health and social care research 2021 [cited 2022 14 July]. Available from: <https://www.nihr.ac.uk/documents/briefing-notes-for-researchers-public-involvement-in-nhs-health-and-social-care-research/27371>.
9. Manafo E, Petermann L, Mason-Lai P, Vandall-Walker V. Patient engagement in Canada: a scoping review of the 'how' and 'what' of patient engagement in health research. *Health Res Policy Syst*. 2018;16(1):5. <http://doi.org/10.1186/s12961-018-0282-4>
10. Aitken M, de St. Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. 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*. 2016;17(1):73. <http://doi.org/10.1186/s12910-016-0153-x>
11. Jones K, Heys S, Thompson R, Griffiths L, Johnson R, Lee A, et al. Linking pre- and post-adoption records for research in anonymised form in a data safe haven: International Journal of Population Data Science. 2020;5(3). <http://doi.org/10.1186/s12910-016-0153-x>
12. Muller SHA, Kalkman S, Van Thiel GJM, Mostert M, Van Delden JJM. The social licence for data-intensive health research: towards co-creation, public value and trust. *BMC Medical Ethics*. 2021;22(1). <http://doi.org/10.1186/s12910-021-00677-5>
13. Jones KH, Heys S, Thompson R, Cross L, Ford D. Public involvement & engagement in the work of a data safe haven: A case study of the SAIL Databank. *International Journal of Population Data Science*. 2020;5(3). <http://doi.org/10.23889/IJPDS.V5I3.1371>
14. Jones KH, Ford DV, Thompson S, Lyons RA. A profile of the SAIL databank on the UK secure research platform. *International Journal of Population Data Science*. 2019;4(2). <http://doi.org/10.23889/ijpds.v4i2.1134>
15. Jones KH, McNerney CL, Ford DV. Involving consumers in the work of a data linkage research unit. *International Journal of Consumer Studies*. 2014;38(1):45-51. <https://doi.org/10.1111/ijcs.12062>
16. Teng J, Bentley C, Burgess MM, O'Doherty KC, Mcgrail KM. Sharing linked data sets for research: results from a deliberative public engagement event in British Columbia, Canada. *International Journal of Population Data Science*. 2019;4(1). <http://doi.org/10.23889/ijpds.v4i1.1103>
17. Deverka PA, Gilmore D, Richmond J, Smith Z, Mangrum R, Koenig BA, et al. Hopeful and Concerned: Public Input on Building a Trustworthy Medical Information Commons. *J Law Med Ethics*. 2019;47(1):70-87. Epub 2019/04/18. <http://doi.org/10.1177/1073110519840486>. PubMed PMID: 30994071; PubMed Central PMCID: PMC6730638.
18. Townson J, Davies J, Hurt L, Ashfield-Watt P, Paranjothy S. Developing and evaluating a model of public involvement and engagement embedded

- in a national longitudinal study: HealthWise Wales. *International Journal of Population Data Science*. 2020;5(3). <http://doi.org/10.23889/ijpds.v5i3.1356>
19. Hogg C. Citizens, Consumers and the NHS. Capturing Voices. Basingstoke: Palgrave Macmillan; 2009.
20. Phillimore J. Approaches to health provision in the age of super-diversity: accessing the NHS in Britain's most diverse city. *Critical Social Policy*. 2011;31(1):5-29. <http://doi.org/10.1177/0261018310385437>. PubMed PMID: 104824827. Language: English. Entry Date: 20110429. Revision Date: 20150711. Publication Type: Journal Article.
21. NIHR. Improving inclusion of under-served groups in clinical research: Guidance from INCLUDE project 2020 [cited 2020 5 November]. Available from: <https://www.nihr.ac.uk/documents/improving-inclusion-of-under-served-groups-in-clinical-research-guidance-from-include-project/25435>.
22. Nelson E, Burns F. Impact through Engagement. Co-production of administrative data research and the approach of the Administrative Data Research Centre Northern Ireland. *International Journal of Population Data Science*. 2020;5(3). <http://doi.org/10.1177/0261018310385437>
23. Ocloo J, Matthews R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Quality & Safety*. 2016;25(8):626–32. <http://doi.org/10.1136/bmjqs-2015-004839>
24. IPFCC. Diverse Voices Matter: Improving Diversity in Patient and Family Advisory Councils. 2018.
25. Boote J, Wong R, Booth A. '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*. 2015;18(1):44–57. <http://doi.org/10.1111/hex.12007>
26. Jewell A, Pritchard M, Barrett K, Green P, Markham S, McKenzie S, et al. 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*. 2019;5(1). <http://doi.org/10.1186/s40900-019-0152-4>
27. Turk A, McPherson A, Boylan A-M, Locock L. Researcher's Guide to Patient and Public Involvement 2020 [cited 2021 12 January]. Available from: <https://oxfordbrc.nihr.ac.uk/wp-content/uploads/2017/03/A-Researchers-Guide-to-PPI.pdf>.
28. Tarpey M. Why people get involved in health and social care research: a working paper 2006 [cited 2021 13 January]. Available from: <https://www.invo.org.uk/wp-content/uploads/2012/01/whypeoplegetinvolvedinresearch2006.pdf>.
29. Boden C, Edmonds AM, Porter T, Bath B, Dunn K, Gerrard A, et al. Patient partners' perspectives of meaningful engagement in synthesis reviews: A patient-oriented rapid review. *Health Expectations*. 2021. <http://doi.org/10.1111/hex.13279>
30. Michie S, Van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*. 2011;6(1):42. <http://doi.org/10.1186/1748-5908-6-42>
31. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of behaviour change theories: Silverback publishing; 2014.
32. Barker F, Atkins L, De Lusignan S. 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*. 2016;55(sup3):S90–S8. <http://doi.org/10.3109/14992027.2015.1120894>
33. Alexander KE, Brijnath B, Mazza D. Barriers and enablers to delivery of the Healthy Kids Check: an analysis informed by the Theoretical Domains Framework and COM-B model. *Implementation Science*. 2014;9(1):60. <http://doi.org/10.1186/1748-5908-9-60>
34. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nature Human Behaviour*. 2020;4(5):451–9. <http://doi.org/10.1038/s41562-020-0887-9>
35. Rubinstein H, Marcu A, Yardley L, Michie S. Public preferences for vaccination and antiviral medicines under different pandemic flu outbreak scenarios. *BMC Public Health*. 2015;15(1):190. <http://doi.org/10.1186/s12889-015-1541-8>
36. Willmott TJ, Pang B, Rundle-Thiele S. Capability, opportunity, and motivation: an across contexts empirical examination of the COM-B model. *BMC Public Health*. 2021;21(1). <http://doi.org/10.1186/s12889-021-11019-w>
37. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research. *Academic Medicine*. 2014;89(9):1245–51. <http://doi.org/10.1097/acm.0000000000000388>
38. Vargas-Silva C, Rienzo C. Briefing: Migrants in the UK: An overview. COMPAS, 2022.
39. Rzepnikowska A. Racism and xenophobia experienced by Polish migrants in the UK before and after Brexit vote. *Journal of Ethnic and Migration Studies*. 2018:1-17. <http://doi.org/10.1080/1369183X.2018.1451308>
40. Teodorowski P, Woods R, Czarnecka M, Kennedy C. Brexit, acculturative stress and mental health among EU citizens in Scotland. *Population, Space and Place*. 2021. <http://doi.org/10.1002/psp.2436>

41. Fischbacher CM, Hunt S, Alexander L. How physically active are South Asians in the United Kingdom? A literature review. *Journal of Public Health*. 2004;26(3):250–8. <http://doi.org/10.1093/pubmed/fdh158>
42. Ranasinghe CD, Ranasinghe P, Jayawardena R, Misra A. Physical activity patterns among South-Asian adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*. 2013;10(1):116. <http://doi.org/10.1186/1479-5868-10-116>
43. Kakde S, Bhopal RS, Jones CM. A systematic review on the social context of smokeless tobacco use in the South Asian population: Implications for public health. *Public Health*. 2012;126(8):635–45. <https://doi.org/10.1016/j.puhe.2012.05.002>
44. Prajapati R, Liebling H. 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*. 2022;9(2):598–619. <http://doi.org/10.1007/s40615-021-00993-x>
45. Lobe B, Morgan D, Hoffman KA. Qualitative Data Collection in an Era of Social Distancing. *International Journal of Qualitative Methods*. 2020;19:160940692093787. <http://doi.org/10.1177/1609406920937875>
46. Squires A. Language barriers and qualitative nursing research: methodological considerations. *International Nursing Review*. 2008;55(3):265–73. <http://doi.org/10.1111/j.1466-7657.2008.00652.x>
47. Garfield S, Jheeta S, Husson F, Jacklin A, Bischler A, Norton C, et al. Lay involvement in the analysis of qualitative data in health services research: a descriptive study. *Research Involvement and Engagement*. 2016;2(1). <http://doi.org/10.1186/s40900-016-0041-z>
48. Hemming L, Pratt D, Bhatti P, Shaw J, Haddock G. Involving an individual with lived-experience in a co-analysis of qualitative data. *Health Expectations*. 2021. <http://doi.org/10.1111/hex.13188>
49. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77–101. <http://doi.org/10.1191/1478088706qp063oa>
50. Braun V, Clarke V. *Thematic Analysis: A Practical Guide*. London: Sage; 2021.
51. Braun V, Clarke V. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*. 2019;11(4):589–97. <http://doi.org/10.1080/2159676X.2019.1628806>
52. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*. 2006;18(1):59–82. <http://doi.org/10.1177/1525822X05279903>
53. Salamońska J, Lesińska M, Kloc-Nowak W. Polish Migrants in Ireland and Their Political (Dis) engagement in Transnational Space. *Central and Eastern European Migration Review*. 2021;10(2):49–69.
54. Teodorowski P, Rodgers SE, Fleming K, Tahir N, Ahmed S, Frith L. ‘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*. 2023. <https://doi.org/10.1111/hex.13713>
55. Laird Y, Manner J, Baldwin L, Hunter R, Mcateer J, Rodgers S, et al. Stakeholders’ experiences of the public health research process: time to change the system? *Health Research Policy and Systems*. 2020;18(1). <http://doi.org/10.1186/s12961-020-00599-5>
56. Grotz J, Ledgard M, Poland F. *Patient and Public Involvement in Health and Social Care Research*: Springer International Publishing; 2020 2020-01-01.
57. Brady L-M. Rhetoric to reality: the need for a new approach. In: Brady L-M, editor. *Embedding Young People’s Participation in Health Services New Approaches*. Bristol: Policy Press; 2020. p. 249–66.
58. Boylan AM, Locock L, Thomson R, Staniszevska S. “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*. 2019;22(4):721–30. <http://doi.org/10.1111/hex.12883>
59. Waing E. Trust, security and public interest: Striking the balance. *International Journal of Population Data Science*. 2020;5(3). <http://doi.org/10.23889/ijpds.v5i3.1368>
60. Moody G, Cannings-John R, Hood K, Robling M. Attitudes towards the collection and linkage of maltreatment data for research: A qualitative study. *International Journal of Population Data Science*. 2022;7(1). <http://doi.org/10.23889/ijpds.v7i1.1693>
61. Bailey Wilson B, Twins B, Wilkinson-Salamea C, Raidos D, Imafidon K, McGarry N, et al. 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, 2022.
62. Mouton Dorey C, Baumann H, Biller-Andorno N. Patient data and patient rights: Swiss healthcare stakeholders’ ethical awareness regarding large patient data sets - a qualitative study. *BMC medical ethics*. 2018;19(1):20-. <http://doi.org/10.1186/s12910-018-0261-x>. PubMed PMID: 29514635.
63. Kirkham EJ, Crompton CJ, Iveson MH, Beange I, McIntosh AM, Fletcher-Watson S. Co-development of a Best Practice Checklist for Mental Health Data Science: A Delphi Study. *Frontiers in Psychiatry*. 2021;12. <http://doi.org/10.3389/fpsy.2021.643914>

Appendix 1

Standards for Reporting Qualitative Research (SRQR)*

<http://www.equator-network.org/reporting-guidelines/srqr/>

	Page/line no(s).
Title and abstract	Page 1
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 1
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 1
Introduction	
Problem formulation - Description and significance of the problem/phenomenon studied; review of relevant theory and empirical work; problem statement	Pages 2-3
Purpose or research question - Purpose of the study and specific objectives or questions	Page 3
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**	Pages 3-4
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 4
Context - Setting/site and salient contextual factors; rationale**	Pages 3-4
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 3
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 4
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**	Pages 3-4
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	Pages 3-4
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 3
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 4
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 4
Techniques to enhance trustworthiness - Techniques to enhance trustworthiness and credibility of data analysis (e.g., member checking, audit trail, triangulation); rationale**	Page 4
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 4
Links to empirical data - Evidence (e.g., quotes, field notes, text excerpts, photographs) to substantiate analytic findings	Pages 4-9

Continued

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	Pages 9–10
	Limitations - Trustworthiness and limitations of findings	Page 10
Other	Conflicts of interest - Potential sources of influence or perceived influence on study conduct and conclusions; how these were managed	Page 11
	Funding – Sources of funding and other support; role of funders in data collection, interpretation, and reporting	Page 11
	*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.	
	<u>Reference:</u> 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	



Appendix 2

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 a description of what big data is.

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?

