
Review

Public Involvement and Engagement in Big Data Research: Scoping Review

Piotr Teodorowski¹, PhD; Elisa Jones², MA; Naheed Tahir³; Saiqa Ahmed³, MSc; Sarah E Rodgers², PhD; Lucy Frith⁴, PhD

¹Faculty of Health Sciences and Sport, University of Stirling, Stirling, United Kingdom

²Department of Public Health, Policy & Systems, University of Liverpool, Liverpool, United Kingdom

³National Institute for Health and Care Research Applied Research Collaboration North West Coast, Liverpool, United Kingdom

⁴Centre for Social Ethics and Policy, University of Manchester, Manchester, United Kingdom

Corresponding Author:

Piotr Teodorowski, PhD

Faculty of Health Sciences and Sport

University of Stirling

Pathfoot Building

Stirling, FK9 4LA

United Kingdom

Phone: 44 1786466362

Email: piotr.teodorowski@stir.ac.uk

Abstract

Background: The success of big data initiatives depends on public support. Public involvement and engagement could be a way of establishing public support for big data research.

Objective: This review aims to synthesize the evidence on public involvement and engagement in big data research.

Methods: This scoping review mapped the current evidence on public involvement and engagement activities in big data research. We searched 5 electronic databases, followed by additional manual searches of Google Scholar and gray literature. In total, 2 public contributors were involved at all stages of the review.

Results: A total of 53 papers were included in the scoping review. The review showed the ways in which the public could be involved and engaged in big data research. The papers discussed a broad range of involvement activities, who could be involved or engaged, and the importance of the context in which public involvement and engagement occur. The findings show how public involvement, engagement, and consultation could be delivered in big data research. Furthermore, the review provides examples of potential outcomes that were produced by involving and engaging the public in big data research.

Conclusions: This review provides an overview of the current evidence on public involvement and engagement in big data research. While the evidence is mostly derived from discussion papers, it is still valuable in illustrating how public involvement and engagement in big data research can be implemented and what outcomes they may yield. Further research and evaluation of public involvement and engagement in big data research are needed to better understand how to effectively involve and engage the public in big data research.

International Registered Report Identifier (IRRID): RR2-<https://doi.org/10.1136/bmjopen-2021-050167>

(*J Particip Med* 2024;16:e56673) doi: [10.2196/56673](https://doi.org/10.2196/56673)

KEYWORDS

patient and public involvement; PPI; involvement; engagement; big data; data science; patient engagement; co-design; coproduction

Introduction

Background

The growth of big data allows researchers to use and link large, multisource health data sets for research. Big data is still an

evolving field [1], and disagreements remain on precisely what the term stands for in health research [2]. Other terms used include routinely collected data [3] and data-intensive research [1,4]. For clarity, throughout this paper, we will refer broadly to the term big data as it is used in the literature and easily

understood by the public. We follow the definition by Aitken et al [1], recognizing that the main feature of big data is the ability to link large data sets for analysis. They name sources for such data as patient records, administrative, registry biobanking, social media, and digital application data. Big data research in health can be used for multiple purposes with the aim of improving health care services and reducing health inequalities [5,6]. These include service management, evaluation or audit of services, statistics, and exploring connections between health and non-health-related outcomes [1]. Often, these purposes differ from the original intent of data collection (eg, health care or statistical purposes). In other words, big data is often used for secondary research purposes.

Big data research offers new opportunities for academics. However, reusing big data for research faces ethical challenges [7]. Previous big data initiatives suggest that the public must have confidence that their data will be used in an acceptable way if they are going to be supportive of big data research [8]. This means moving outside what is legally required and establishing a social license for research [9]. Carter et al [9] proposed 3 conditions for establishing a social license for big data research. First, reciprocity is essential, as there is a need for 2-way communication and improving public awareness of big data research as well as improving researchers' understanding of the public's concerns and expectations. A lack of transparency could make it challenging to secure public trust [10], and the public has a right to be informed about the progress of the research [11]. Second, the process should empower, not disempower, the public; in big data research, this could include members of the public involved in the governance of data linkage and the design of big data projects. Third, big data research should benefit the public; thus, researchers need to understand what the public might perceive as public benefit.

Public involvement and engagement could be used to bridge the gap between researchers and the public's understandings of the benefits of big data research [12]. There is evidence in the literature (outside big data) that public involvement can provide legitimacy for research [13]. Public contributors could be a part of the process of creating research norms for big data research [14]. Research norms consist of governance and regulation that could guide research. These might not be popular among some academics, but they could help secure a social license for research [15]. Aitken et al [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 can be meaningfully involved in big data research projects [16-18]. However, there is a need to understand how public involvement and engagement takes place in big data research comprehensively.

Objectives

Previous reviews have examined literature around public trust and attitudes toward big data research [19-22]. Despite public involvement and engagement being seen as one of the ways to improve public trust, as far as we are aware, there have not been any previous reviews exploring public involvement and

engagement in big data research and there have not been any reviews registered on the PROSPERO and Cochrane databases. Therefore, this review aimed to synthesize what is known about public involvement and engagement in big data research. Using scoping review methodology [23-25], we mapped key issues in the research to find evidence of how public involvement and engagement were carried out in big data research. Understanding how to involve and engage the public in big data research could be used to formulate guidance for researchers and policy makers on how to do this effectively, as there are field-related challenges, especially regarding the abstraction and complexity of big data [26].

Methods

Overview

The protocol for this scoping stage review was published previously [27]. The protocol outlines the parameters of the review and provides a justification and explanation of all the methodological steps and decisions taken. To ensure rigor further, we used the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [28] and reported it as [Multimedia Appendix 1](#).

Defining Public Involvement

In the literature, the terms involvement, engagement, and participation are used interchangeably, but these do not always have the same meaning [29,30]. This makes research and discussion about public involvement challenging, as it can be difficult to identify papers for review [31-33]. Hence, there is growing recognition that more consistent terminology is needed [13]. The diversity of types of involvement can be seen in the ladder by Arnstein [34] that determines types of involvement by constructing a typology based on the amount of power given to the public. It identifies from the bottom (lowest extent of people's influence) to the top (highest extent of people's influence) the following steps: therapy manipulation, nonparticipation, informing, consultation, placation, partnership, delegation, and full citizen control. The author herself called the ladder "provocative." One of the health-specific definitions of public involvement has been developed by INVOLVE [35]. It has been used broadly by funders and researchers and embedded in the public involvement reporting checklist [33]. It offers a nuanced perspective on 3 types of activities: involvement, engagement, and consultation, which researchers can use when working with members of the public. One is not better than the other, but rather, each offers a different approach. INVOLVE defines involvement as research carried out with or by members of the public rather than to, about, or for them. This recognizes shared ownership of research with members of the public. Engagement is providing information about big data research and disseminating it to the public. Consultation happens when the research is discussed with the public, but there is no shared ownership. Thus, engagement and consultation are "to," "about," or "for" rather than "with" or "by" them. However, these activities can provide an understanding of the public views.

Owing to the diversity of definitions of public involvement and engagement used in the literature, we mapped all included

papers using the INVOLVE definition, identifying whether they were involvement, engagement, or consultation.

Public Involvement in the Review

Public involvement in reviews can improve their quality by contributing to defining the scope, appraising the papers, and interpreting results [36,37]. In total, 2 public contributors (SA and NT) were involved in the review from the initial design stage and contributed at each stage (screening, data extraction, and analysis). They are both experienced public contributors and previously copublished papers around public involvement and engagement in big data research. SA and NT ensured the relevance of review results to the public. This was achieved by relating results to their experience as public contributors in other research projects. The details of the involvement process and what was put in place to support them (eg, training) are reported elsewhere.

Searches

Following the search strategy developed with the support of a university librarian, the CINAHL, Health Research Premium Collection, PubMed, Scopus, and Web of Science databases were searched for papers in September 2021. The search strategy, as published in the protocol paper, is included in [Multimedia Appendix 2](#). The search covered papers published after 2010 until the search completion in September 2021. Additional manual searches were conducted. These included the screening of the first 100 results from a Google Scholar search, journals that aim to publish public involvement research (*BMC Research Involvement and Engagement* and *Health Expectations*) or had special editions on public involvement in big data (*International Journal of Population Data Science*), and gray literature (the first 100 results from the Patient Outcome Research Institute database were screened). A call for potential papers to be included was posted on X (previously known as Twitter) to reach experts in the field.

Inclusion Criteria

The review included papers that met the following criteria: (1) discussed public involvement or engagement in big data research (those that appeared more as consultations were not excluded, but a note was taken of this), (2) focused on patient- or health-related research, and (3) were published in English. All study designs and nonempirical discussion papers were included.

Screening and Study Selection

PT took the lead by screening all papers. SA, NT, and EJ jointly screened at least a random 20% of papers at each stage (title, abstract, and full paper). Any discrepancies were discussed by the research team. The reasons for exclusions at a full paper stage were recorded and reported in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.

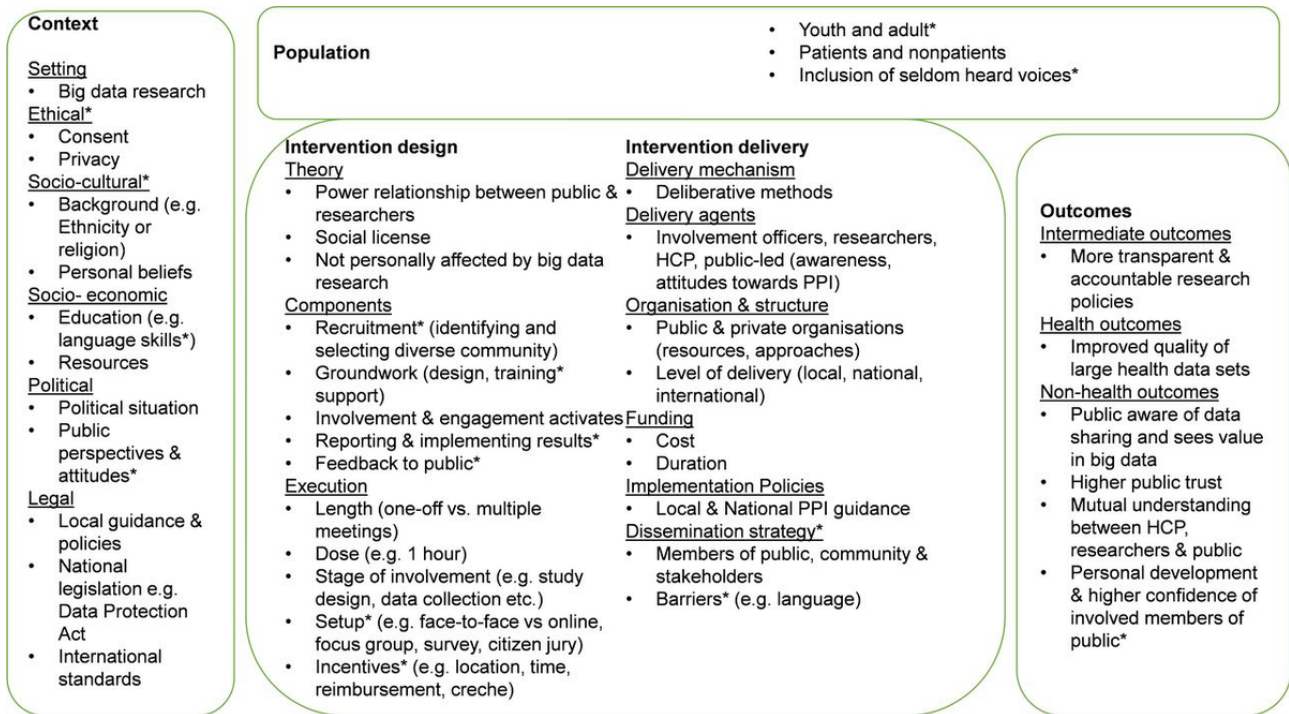
Data Extraction

The data extraction form development was iterative and tested by the whole research team. The final data extraction form is available in [Multimedia Appendix 3](#). PT extracted data from all papers in the first instance. Then, all extraction was double checked by the rest of the research team, thus ensuring each paper was considered by 2 researchers. The research team met regularly to discuss any discrepancies and discuss initial findings. PT organized the extracted data in a descriptive and narrative way under key headings based on the data extraction form. This was discussed with the research team.

Analysis

The analysis was supported by a prior system logic model that we published in the protocol paper ([Figure 1 \[27\]](#)). It was initially developed by a preliminary scoping of the literature, research team discussion, and input from the public contributors. The logic model assisted us in identifying relevant elements of public involvement and engagement in big data research. We mapped our findings under the model and present them using headings from the logic model.

Figure 1. System logic model of public involvement and engagement in big data research (reproduced from the study by Teodorowski et al). HCP: health care provider; PPI: public and patient involvement.



Results

Overview

The database searches produced 4054 papers. Additional manual searches added a further 11 papers. After the removal of duplicates, 3540 articles were screened for inclusion in the

review. A total of 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 2 \[38,39\]](#) shows the PRISMA flowchart of the screening process. We first discuss the study characteristics and thereafter present findings as mapped under the revised system logic model ([Figure 3 \[27\]](#)).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. PPIE: patient and public involvement and engagement.

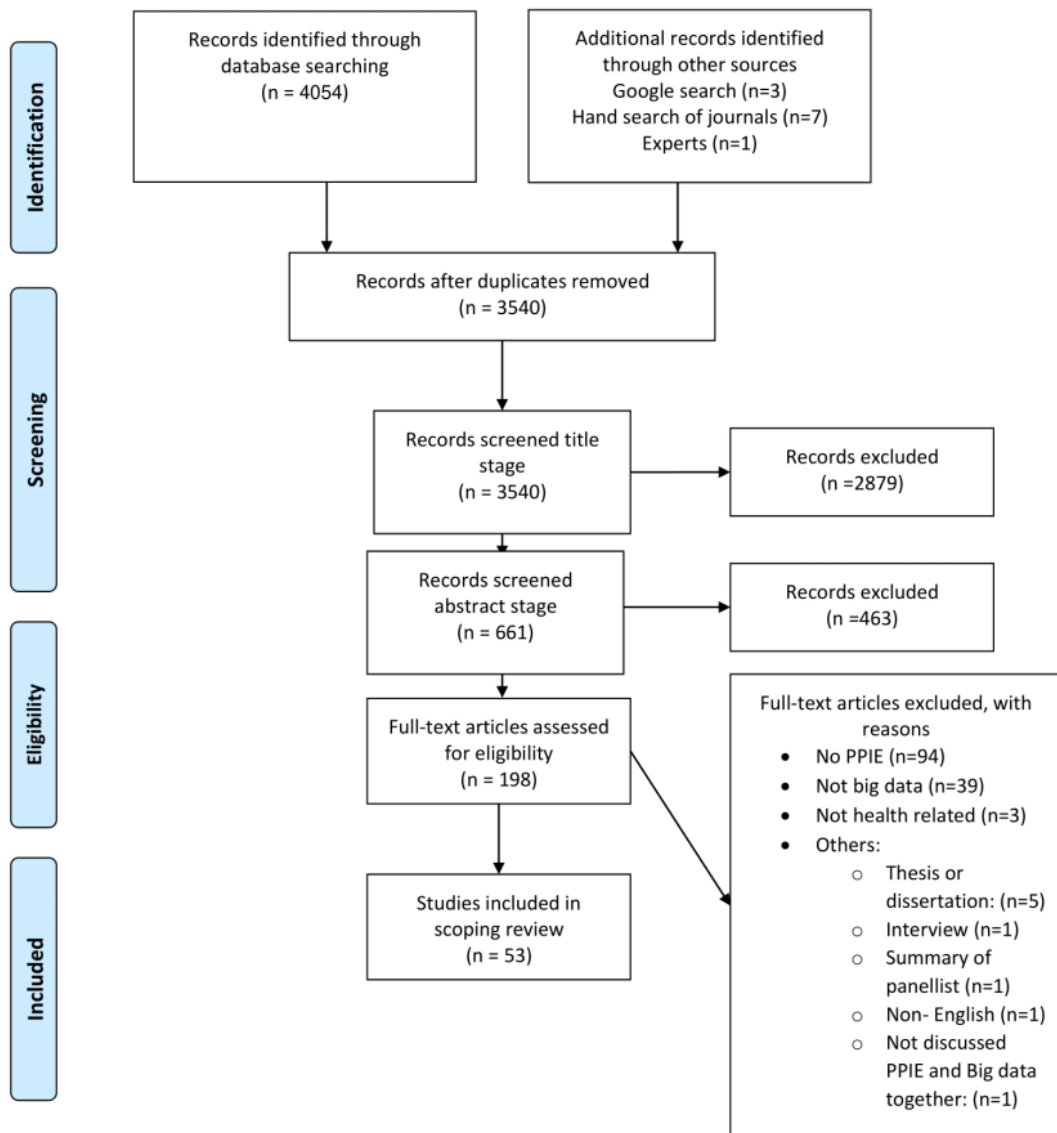
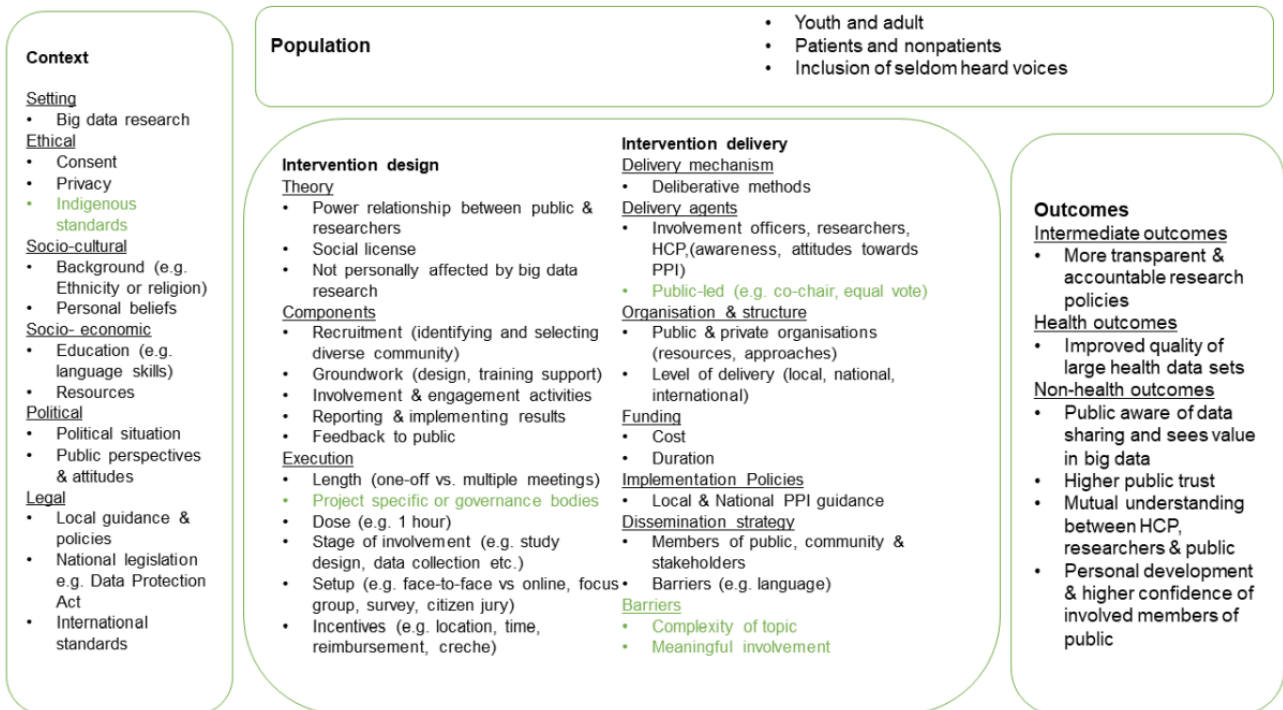


Figure 3. The updated a priori system logic model of public involvement and engagement in big data research (adapted from the study by Teodorowski et al). Green color is used to record new aspects of the model based on the review. HCP: health care provider; PPI: public and patient involvement.



Study Characteristics

The most prevalent type of papers were discussion papers (nonempirical, including conceptual or ethical papers; 28/53, 53%), followed by review papers (5/53, 9%); qualitative study design (5/53, 9%); opinion, letter, commentary, or editorial (4/53, 8%); evaluation (3/53, 6%); protocol (2/53, 4%); ethnographic or descriptive case study (2/53, 4%); public deliberations (1/53, 2%); action research (1/53, 2%); quantitative (1/53, 2%); and mixed methods (1/53, 2%). The papers were from the United Kingdom (19/53, 36%), the United States

(10/53, 19%), Canada (7/53, 13%), New Zealand (3/53, 6%), the Netherlands (1/53, 2%), Portugal (1/53, 2%), France (1/53, 2%), South Africa (1/53, 2%), Australia (1/53, 2%), Germany (1/53, 2%), and Africa (1/53, 2%). In total, 12 papers did not specify a geographical location, and some papers included more than one. The most prevalent type of involvement and engagement activities carried out with the public (following INVOLVE definitions) were involvement (45/53, 85%), followed by engagement (25/53, 47%) and consultation (7/53, 13%). Some papers discussed >1 type of activity. Table 1 presents the characteristics of the included papers.

Table 1. Summary of the included papers in the scoping review.

Study; year	Design	Location	Demographics to involve and engage	Types of activities	Area of interest
Baart and Abma [40], 2010	Action research	Netherlands	Not specified	Involvement and engagement	Involvement in psychiatric genomics research
Ballantyne and Style [41], 2017	Discussion	New Zealand	Lay, gender, and Māori representation	Involvement and engagement	Expert health data research ethics committee
Ballantyne and Stewart [42], 2019	Discussion	United Kingdom	Affected group; priority is given to patient groups considered vulnerable	Involvement and engagement	Public and private sectors collaborate to share, analyze, and use biomedical big data
Beyer et al [43], 2010	Qualitative	United States	Caucasian, Hispanic, Taidam or Lao; represented various education, income, and other characteristics	Involvement and consultation	Geocoded health information and experiential geographical information in a GIS ^a environment
Bharti et al [44], 2021	Discussion	United Kingdom	Not specified	Engagement	Securing public trust and the importance of public engagement
Bot et al [45], 2019	Discussion	United States	Underrepresented populations	Involvement	Decentralization of governance
Coulter [46], 2021	Editorial	United Kingdom	General public	Involvement	National Health Services Digital plans to update its systems from patient data from general practitioner records
Dankar et al [47], 2018	Discussion	N/A ^b	Not specified	Engagement	Data governance in population genome projects
de Freitas et al [48], 2021	Protocol	Portugal	Patients and informal carers	Involvement	Coproduction of a people-centered model for the public in decision-making processes about data reuse
Deverka et al [49], 2019	Public deliberations	United States	Diverse geographic and individuals with chronic illness	Involvement and consultation	Recommendations for medical information commons design and management
Duchange et al [50], 2014	Discussion	France (European Union project)	Representatives of patient organizations	Involvement, engagement, and consultation	Ethics committee
Erikainen et al [51], 2020	Qualitative	United Kingdom	Not specified	Involvement	Governance of population-level biomedical research
Evans et al [52], 2020	Qualitative	United States	Individuals with OUD ^c and their families	Involvement and engagement	Reuse of big data on opioid use
Fernando et al [53], 2019	Letter	South Africa	Traditional community leaders	Involvement and consultation	Data governance model in biobanking and data sharing
Fleurence et al [54], 2014	Discussion	United States	Patients	Involvement	National research network (PCORnet)
Funnell et al [55], 2020	Discussion	Canada	Indigenous communities	Involvement	Community-based participatory research methods in a project using previously collected data to examine end-of-life health care
Gallier et al [56], 2021	Discussion	United Kingdom	Not specified	Involvement and engagement	PIONEER infrastructure and data access processes
Goytia et al [57], 2018	Qualitative	United States	Patients	Involvement and engagement	Views on big data research
Henare et al [58], 2019	Opinion	New Zealand	Indigenous people	Involvement and engagement	Road map for neuroendocrine tumor research to reflect the values of Indigenous people
Hudson et al [59], 2020	Discussion	N/A	Indigenous population	Involvement	Indigenous communities' views on the sharing of genomic data
Hurt et al [60], 2019	Discussion	United Kingdom	Not specified	Involvement and engagement	Design of HealthWise Wales

Study; year	Design	Location	Demographics to involve and engage	Types of activities	Area of interest
Jewell et al [61], 2019	Evaluation	United Kingdom	Service users and carers	Involvement	Advisory group
Jones et al [18], 2013	Evaluation	United Kingdom	Consumers; at least 1 representative from an ethnic minority group	Involvement	Consumer panel
Jones et al [17], 2019	Discussion	United Kingdom	Not specified	Involvement and engagement	SAIL Databank
Jones et al, 2020 [16]	Evaluation	United Kingdom	Inclusive of all ages, ethnic groups, cultures, socioeconomic levels, lifestyles, and other definable interests	Involvement and engagement	SAIL Databank and related population data science initiatives
Kalkman et al [62], 2019	Systematic review	N/A	N/A	Involvement and engagement	Ethical guidelines for principles and norms pertaining to data sharing
Kirkham et al [63], 2021	Qualitative	N/A	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 [64], 2020	Scoping review	N/A	N/A	Involvement	Reporting the impact of public involvement in biobanks
Manrique de Lara and Peláez-Ballestas [65], 2020	Narrative review	N/A	N/A	Involvement and engagement	Bioethical perspectives of big data
Milne et al [66], 2021	Discussion	United States and North America	Not specified	Involvement	Data trust model in the governance of biobanks
Milne and Brayne [67], 2020	Discussion	N/A	Not specified	Involvement	Data governance in dementia
Mourby et al [68], 2019	Discussion	United Kingdom	Not specified	Involvement and engagement	Obstacles preventing data linkage research from reaching its full potential
Murtagh et al [69], 2018	Ethnographic case study	United Kingdom	Participants of genomic studies	Involvement and engagement	Foundational principles of data sharing infrastructure
Nelson and Burns [70], 2020	Discussion	United Kingdom	Most affected communities by the research	Engagement	ADRC NI ^d approach to public engagement
Newburn et al [3], 2020	Discussion	United Kingdom	Service users; 1 activity targeted ethnic minority groups	Involvement and engagement	Service user participation in a data linkage study
Nunn et al [71], 2021	Mixed methods	Australia	Not specified	Involvement	Involvement in genomic research
O'Doherty et al [72], 2011	Discussion	Canada	Groups considered historically disadvantaged	Involvement and engagement	Biobank governance and principles to form governance structures
O'Doherty et al [73], 2021	Commentary	N/A	Not specified	Involvement	Functions of good governance
Ohno-Machado et al [74], 2014	Discussion	United States	Patients	Involvement and consultation	Setting up of the pSCANNER ^e
Omar et al [75], 2020	Discussion	N/A	Not specified	Involvement, engagement, and consultation	European network of excellence for big data in prostate cancer
Paprica et al [76], 2020	Discussion	Canada	Communities facing long-standing inequalities that are affected by the research	Involvement and engagement	Establishment and operation of data trusts

Study; year	Design	Location	Demographics to involve and engage	Types of activities	Area of interest
Patel et al [77], 2021	Quantitative	United Kingdom	Not specified	Involvement	The use of remote consultation and prescribing of psychiatric medications
Pavlenko et al [78], 2020	Systematic review	N/A	N/A	Involvement	Governance in clinical data warehouses internationally
Rowe et al [79], 2021	Discussion	Canada, New Zealand, and United States	Indigenous people	Involvement	Principles for linking Indigenous population data
Shaw et al [11], 2020	Discussion	United States, Canada, and United Kingdom	General public and specific communities (eg, African Americans, Indigenous people, people with disabilities, and people living with homelessness)	Engagement	Social license for big data initiatives
Sleigh and Vayena [80], 2021	Descriptive case study	Germany and United Kingdom	General public	Engagement	Visual public engagement campaigns
Teng et al [81], 2019	Discussion	Canada	Not specified	Involvement	Public deliberation event on the data linkage and reuse for research
Tindana et al [82], 2015	Review	Africa	People affected by the research	Involvement, engagement, and consultation	Community engagement in biomedical and genomic research
Townson et al [83], 2020	Discussion	United Kingdom	Not specified	Involvement and engagement	A model of public involvement and engagement
Vayena and Blasimme [84], 2017	Discussion	N/A	Patients	Involvement	Models of informational control in data-intense health care and clinical research
Weich et al [85], 2018	Protocol	United Kingdom	Mental health users and carers and people with lived experiences; ensure diversity of age, gender, and ethnicity	Involvement	Spatial and temporal variation in the use, effectiveness, and cost of community treatment orders through the analysis of routine administrative data
Willison et al [86], 2019	Discussion	Canada	Patient representatives with diabetes including Francophone, immigrant, and Indigenous populations	Involvement	Governance model for health data repositories
Xafis and Labude [87], 2019	Discussion	N/A	Not specified	Involvement and engagement	Ethics framework for big data in health and research

^aGIS: Geographic Information Systems.

^bN/A: not applicable.

^cOUD: opioid use disorder.

^dADRC NI: Administrative Data Research Centre Northern Ireland.

^epSCANNER: patient-centered Scalable National Network for Effectiveness Research.

Population

The demographics of the public or communities involved and engaged in big data research were diverse. These included patients (including consumers and service users; 12/53, 23%); affected groups or groups considered vulnerable (8/53, 15%); Indigenous communities (6/53, 11%); articles focusing on specific characteristics (eg, gender, age, income, education, or geography; 5/53, 9%); carers (4/53, 8%); the general public (3/53, 6%); ethnic minority groups (3/53, 6%); patient

representative or community leaders (3/53, 6%); and research study participants (1/53, 2%).

Deciding who should be on advisory boards, how they should be selected, and what their role should be remained a challenge for researchers [82]. An important issue was representativeness; advisory boards were unlikely to represent all the public views [66,69,87]. No single committee could represent all communities (because of their diversity) [58,76]. Identifying the relevant communities was seen to be difficult [82]. This created the challenge of ensuring legitimate group representation [72]. Advisory groups often did not reach a broader population [68];

hence, involvement and engagement need to move away from the “usual suspects” [16,18,66,76]. There was the risk that more vocal individuals could dominate the discussion [82]. Public contributors could be chosen arbitrarily, for example, based on personal contracts, and thus, the process might not be transparent to the public [72]. This could lead to involving financially and politically motivated [49] or well-connected contributors [42]. The way to overcome these issues could be to recruit public contributors from the study participants; for example, participants could elect their own representatives or a marketing company could conduct the recruitment [72,81].

Context

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 [59,79]. Researchers should incorporate Indigenous culture, for example, traditional ceremonies, when involving the community [58]. Formalized agreements with Indigenous organizations could improve the relationship with that community [55]. This more nuanced approach to big data research could assist researchers in establishing trust with Indigenous communities rather than merely convincing them that this is the right thing to do [59].

Political situations or public perspectives and attitudes could influence how and why members of the public get involved in big data research. Secrecy could be a challenge [11]. Organizations might not want to share controversial information, and private companies may argue that sharing it might be against their commercial interests [42]. Involvement and engagement could have the potential to improve public trust in big data research but not necessarily in the research institution [51]. There could be historic mistrust from underserved communities, for example, African Americans, Indigenous communities, and people living with homelessness [11]. There was no guarantee that it would always be possible to maintain public trust in big data research [67].

Intervention Design

Theory

Respectful, ongoing, genuine, and nonhierarchical interaction between researchers and the public was seen as necessary to build trust [16,87]. Building a relationship could take time [82]. It included the coownership of research [55] and should concentrate on what the public wants to know [40]. The reciprocal relationship was illustrated by Newburn et al [3], who organized workshops during which they delivered training for members of the public on using social media and research methodology. A clear purpose for the activity leads to realistic expectations [16]. The starting point for involvement might not be about assuming an equal partnership but an exploration of power relationships [40]. Working in smaller groups gave more opportunities for every public contributor to share their opinion [81]. Decisions could be made through consensus [55,86]. However, Ballantyne and Stewart [42] recognized that there would always be disagreements and that all opinions cannot always be acted on; in that case, there might be a need for a clear explanation of why these voices were not included.

Conducting involvement and engagement activities did not mean that public values are incorporated into big data research [72]. Involvement could be tokenistic without effecting real change, but this still could offer some form of legitimacy to researchers and the research [72]. There was a need to ensure a balanced power relationship between public contributors and the research team [48]. When public contributors joined already ongoing research projects, they had limited scope for impact (eg, amendments might not be allowed); thus, their involvement might turn more into consultation [66,82]. Some researchers did not support involvement and would prefer a deficit engagement model where the members of the public were simply informed about the research [40]. Researchers should reflect on how to ensure balance in engagement. It could be about raising awareness of big data research and understanding that it should not be limited to an already agreed outcome but rather an ongoing dialogue [16,17,76]. Public involvement and engagement should take place before any data sharing occurs [11].

Recruitment

Various ways could be used to reach diverse audiences [75,86]. Recruitment of public contributors was mostly through already existing groups such as involvement groups (eg, Jewell et al [61] used an established involvement register that was open for service users and their families or carers), patient organizations [18,61,74,75,85], clinical sites [74], or recruitment via newsletter distributed among study participants [60,83]. Working with intermediaries (eg, charities or community leaders) could improve the reach as they can provide advice about public perspectives or can become gatekeepers [70,82]. Public contributors might be unclear on their role at the beginning [18]. Therefore, clear criteria for the public are needed [66]. Promoting involvement should focus on seeing it as a reciprocal opportunity with benefits for both researchers and public contributors [3]. The recruitment advertisement should include a description of the role and the required skills [61]. The full research protocol with all methodological details should be available on request [3]. There was a perceived need for a transparent process of selecting public contributors to avoid tokenism [49,73]. Candidates could be interviewed to identify individuals with team working skills and the ability to contribute outside their own health situation [18,86], as public contributors’ emotional connection to the research could be both an enabler or a barrier to their involvement [71].

Engagement is about reaching the broader public, especially around dissemination [62,68]. The engagement was mentioned alongside education, as it showed how findings from big data projects were shared with the community [52]. Educating the public could be seen as paternalistic, one directional, and top down; hence, there was a need for 2-way communication [11,86]. Researchers should share any discussion from governance groups with a broader public [11,56]. These could be a brief web-based report of findings and key recommendations [43].

Contribution

Public contributors had various roles in big data research. First, they contributed to specific research projects. In some papers,

the public contributors were involved at all stages, from study design and identifying research questions to analysis and dissemination [3,48,52,53,55,57,61,65,82,85,87]. Public contributors also acted as coinvestigators in big data research projects [3].

The other role was around data governance. Public contributors (or representatives of patient organizations) could be involved in (joint) data governance to ensure that research was done ethically (in terms of public interest and sensitivity risk), for example, by advising, cofinding new solutions, or cocreating guidance and policy [16,18,41,42,45,54,56,58-60,62,64,66-69,72-78,86]. Working with the public could offer a lay perspective and ensure that data access and research were in the public interest, and thus, this was argued to potentially pave the way for establishing public trust [17,18,41,56,60,66,68].

One paper reported that public contributors who were members of governance bodies acted as big data advocates [16]. However, their voice should be of equal value as other stakeholders [49]. For example, if the group felt that a big data project did not have enough public input, they could assign a public contributor to support that particular work [86]. The governance bodies could also assist with engaging the general public (eg, by reviewing lay information) and guide the recruitment of new public contributors [16]. The influence of governance groups differs, and O'Doherty et al [72] recommended flexible governance that could evolve as big data research develops. Some papers argued that a one-size-fits-all solution might never work in big data research or for diverse communities [45,58,68,82]. Embedding involvement in the governance of big data research may require novel solutions [51].

The public should receive understandable and educational information on project outcomes [75]. Engagement activities should be proportional to the nature and size of the project around big data research [42]. Therefore, the way these engagement activities looked differed between the papers that were included. The public could be reached through engagement events [16,65]. Events were held with service users [3]. Researchers attended and supported events, for example, during the colorectal cancer awareness month [43]. Interactive elements (graphics, videos, etc) were used during exhibitions to raise public awareness [80].

The consultation approach consisted of surveys [50,75], informal small group meetings (eg, town hall meetings) [82], or qualitative research that aimed to capture the public perspective before setting up the project using that community data [53]. These included focus groups (eg, exploring patients' approach to patient engagement in governance and prioritizing research questions) and interviews (eg, to understand public views toward privacy) [74].

In-person activities could be time restrictive and cost restrictive for some communities [74]. Public contributors might not be able to attend meetings, sometimes without warning because of personal circumstances (eg, health treatment, work, or family responsibilities) [3,81].

Intervention Delivery

Delivery Mechanism

Involvement around governing big data research could also be conducted as a one-off deliberation event [49,81] or a Delphi study [63]. A one-off deliberation process could be particularly beneficial for contentious issues [73].

Delivery Agents

Governance groups could be chaired or cochaired by a public contributor, and most members of these groups could be members of the public [60,66,74,86]. If there was >1 governance group in the organization, public contributors could sit on different panels [16-18,74]. The public could be a part of the engagement process. Townson et al [83] mentioned the role of "Champions" who promoted studies in general practitioner surgeries, large public events (eg, food festivals) reaching schools, and support events organized by researchers. Another role they had was that of "supports." Supports (similarly, to champions) were 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 [83].

Involvement and engagement should be led by team members experienced in organizing and running these activities [16,48,60,70,76]. Other researchers should dedicate time to these activities (and this time should be embedded in the workload) [16]. Research team members and facilitators should be trained in public involvement [60,81]. Access to specialist training on involvement and engagement should be provided to both staff and the public [16].

Organization and Structure

Using modern technology, researchers could create a registry or website where the public can see who had access to their data and for what purpose or receive newsletters [3,41,47,72]. Newburn et al [3] aimed to share their research on social media (Twitter and Facebook). Nationwide campaigns could explain the benefits of big data research [52,57,80]. This should be done in the language (eg, Indigenous) the public understands [58]. The public could be further reached through patient organizations [3,75], and researchers could share (yearly) updates jointly with them [50].

Funding

Expectations around monetary compensation should be established from the start [82]. These could include reimbursement for time [61,72,81,83], travel [81], and childcare expenses [3]. Researchers should provide lunch [3] and use venues that are easily accessible by public transport [3]. If public contributors are paid equally to professionals in governing bodies, this might improve their involvement [49].

Implementation Policies

A minority of papers directly referred to involvement or engagement guidance. These included the UK National Standards for Public Involvement [16,60,61], National Institute for Health and Care Research (NIHR) definitions of involvement and engagement [3,83], the GRIPP2 (Guidance for Reporting

Involvement of Patients and the Public) checklist [3,61], the consensus statement on public involvement and engagement with data-intensive health research [16], an academic model guiding involvement [40], and local policies or principles [47,79].

Some papers mentioned legal documents to justify involvement and engagement. These include data protection legislation [16,67], government policies [41,45], and legislation or treaties around Indigenous communities' rights [55,58].

Dissemination Strategy

Researchers should communicate clearly, in lay language and without jargon, to ensure transparency [18,49,76]. The examples included jargon-free graphics [80], tailoring academic research to lay audience [40], and postsession informal debrief [69]. 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 [3]. Public contributors should receive training introducing them to big data research [18,48,69,83,86]. The availability of good-quality information on big data underpins meaningful public involvement [75,87]. Explanations could include links to Wikipedia [83]. Researchers should send information before activities to give people time to reflect on it [83]. Public contributors might need extra time to consider their responses [16].

Barriers

Meaningfully including public contributors in the governance of big data projects could be challenging. Big data could be a complex topic, and it is difficult to find, involve, and engage public contributors with sufficient big data expertise [18,40,47,49,52,57,65]. Potential contributors might feel apprehensive about contributing to complex research if they do not understand the technical jargon [16,42]. This could be further compounded by language and cultural barriers between researchers and the public [82]. Public contributors should be offered training and additional support as required, especially with complicated topics [61,83]. Support needs to be person-centered and based on each individual's skills and experience [83]. These could include short lectures, group discussions, and opportunities to ask questions [61,66]. For example, Teng et al [81] 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 [69]. Kirkham et al [63] included public contributors with big data research experience. Still, they recognize that people with a better understanding of big data might have different views than the general public.

Public involvement should be a meaningful process. Included papers suggested several ways to ensure that members of the public would feel comfortable and able to share their views. Before meeting other stakeholders, public contributors could meet first together [48]. When commenting on a new aspect of research, public contributors were invited to comment first [86]. Some papers described the beginning of the involvement process [40,81]. In the study by Teng et al [81], during the first day of

activities, presentations were made to provide some background on big data research for public contributors. These were from the perspective of patients and seldom-heard communities. These presentations were not neutral but opinionated to show diverse views on big data research.

Outcomes

Some included papers in the review claimed that involvement and engagement should have clear outcomes. First, it could identify gaps in knowledge and priorities for research [70]. Second, it could align researchers' and institutional perspectives of public interest with public views [44], for example, by bringing together charity workers, service providers, elected politicians, and members of the public [54,70]. Third, public contributors involved in governing bodies could have the effect of improving trust and accountability [84]. Fourth, improving public awareness of big data might democratize health research [62]. For example, Vayena and Blasimme [84] argued further that blending citizen science and participatory models could offer more democracy in governance.

However, measuring the impact of involvement and engagement in big data research was challenging [3,64,82,83]. A scoping review by Luna Puerta et al [64] recognized that there was no consensus about the objectives of public involvement in big data research, which undermines the ability to measure impact. Another review by Tindana et al [82] found that the papers included in their review on community engagement did not evaluate the effectiveness of engagement activities.

Engagement through genuine public debate could help demonstrate that the public sector could be a trustworthy steward of patient data [42]. This should include any negative comments toward the initiative; these should be publicly shared, and justification should be provided as to why their feedback was not implemented [42]. Dankar et al [47], when discussing biomedical databases, suggested that sharing research findings should include reaching individuals with personalized research results; these need to be valuable and benefit individuals (eg, they could go for health tests or make life changes that improve their health).

Discussion

Principal Findings

This scoping review provides an overview of how public involvement and engagement have been used in big data research or how it has been argued that it could be applied. This is the first review exploring this issue. The review has shown that the public can and, many articles argue, 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. Those responsible for working with the public should consider what type of activities are most relevant to their work and should use multiple approaches (involvement, engagement, and consultations) to reach different communities. Some papers suggested using modern technology when engaging the public (eg, through a website or digital newsletter). However, most included papers were not primary studies.

The review indicates that many believe that public involvement and engagement have the potential to improve public trust and accountability for big data initiatives. However, there is limited literature on how public involvement and engagement might influence it. Future research should attempt to measure the impact of involvement and engagement in securing social license for big data research with the broader public. The initial step to improve this situation could be to ensure reporting by using standardized reporting guidance for public involvement, such as GRIPP2 [88].

References to public involvement and engagement guidance or legal documents in the included papers were limited. The consensus statement on public involvement and engagement with data-intensive health research [1] is relatively new. However, INVOLVE (now incorporated into the NIHR) has been active in the United Kingdom 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. However, more big data-specific guidance is being developed by the Public Engagement in Data Research Initiative in the United Kingdom [89].

The findings of this review indicate that some challenges are particularly relevant for involvement and engagement in big data research. However, the review has also shown that public involvement and engagement in big data research are not dissimilar to other research fields, as they share aspects of involving and engaging the public, such as working with seldom-heard communities and addressing power balance. This suggests that big data researchers could also use generic public involvement resources, such as the National Standards for Public Involvement in the United Kingdom [90].

The main challenge is that big data research is a complex topic. It might not be easy to explain it briefly (or in accessible language) to potential public contributors or the public. The papers offered some suggestions on how these barriers could be overcome. Researchers need to ensure that they allocate sufficient time and resources when discussing big data research with members of the public. This finding aligns with another review that examined patient involvement in cancer research, where the authors identified time-consuming involvement as a primary challenge in that context [91]. This review suggests that involving and engaging the public in big data research might be even more time consuming than in other fields. If these challenges are overcome, there is a higher chance that involvement and engagement in big data research is not tokenistic, but this might mean additional time and financial

resources. Researchers should budget for these resources as they design any involvement or engagement activities. However, they should be supported to do it by research institutions and funders.

Bailey et al [92] reported that Black and South Asian communities in the United Kingdom have less trust in the health system, and because of this, there might be concerns within these groups about how the public bodies use their data. Researchers need to recognize how trust and attitudes toward big data research could influence public involvement and engagement. This review has offered some indication of how to achieve this from the literature that explored working with Indigenous communities, such as recognizing communities' beliefs and way of life.

The protocol that this review was based on presented the priori system logic model for public involvement and engagement in big data research [27]. On the basis of the review findings, the model was revised. Within the context section, Indigenous standards were added to recognize that big data research needs to consider the perspective and views of Indigenous communities that might differ from previous dominant perspectives. In the intervention theory section, the execution of involvement activities could be divided into project-specific aspects (eg, focusing on 1 big data research project) and governance bodies that look into granting approvals into data linkage (for other projects). These 2 purposes might influence how researchers involve and engage the public. In intervention delivery, the bullet point around public-led activities was added, as some papers suggested that it was important to ensure that the public voice is equivalent to professionals' views during voting and should have equal or even more influence (eg, by cochairing meetings or being coinvestigators). Furthermore, a new bullet point was added in intervention delivery to recognize big data-specific barriers, especially jargon, and how complex big data research could be to members of the public.

Most of the elements included in the model were discussed in the included papers. The only exception is that it does not reflect on the involvement and engagement of people who are not personally affected by big data research (or do not perceive themselves as such). The coverage of most of the issues raised in the papers for involvement and engagement in big data research suggests that the logic model could support researchers who intend to design and deliver these activities to the public.

Textbox 1 provides a summary of the key recommendations around public involvement and engagement in big data research based on the review findings.

Textbox 1. Key recommendations around public involvement and engagement in big data research.

- Ensure that complex and abstract language is explained in lay terms and is understandable to members of the public.
- As public involvement and engagement in big data research might require additional time and resources, these should be planned and budgeted in research plans.
- Trust and public attitudes could influence how and if members of the public get involved in big data research. Public involvement and engagement activities targeting seldom-heard communities should recognize the cultural beliefs held by these groups.
- Following big data research standards could provide researchers with more specific guidance for working with members of the public. These should be used alongside already existing generic guidance.
- Capture and evaluate the impact of public involvement and engagement activities in big data research.

Limitations

The first limitation is the use of terminology. The review explored public involvement and engagement in big data research. These terms are used in different ways by researchers. This parallels the experience of Brett et al [93] in their review, where they found that the variability in wording used to describe involvement complicated literature searching. The search strategy was developed with an experienced librarian and included additional manual searches. However, this did not guarantee that all relevant papers were included. This could have influenced the search results, as potentially some relevant papers might not have been picked up by the search as the authors used different terms. The second limitation was that only information included in the papers was extracted. The authors of included papers were not approached for more details. As academic papers have a word limit, it is possible that some additional information about involvement and engagement may have not been included in the published paper. In contrast to the initial plan, the references of included papers were not screened for potential inclusion. This was because screening of

references of included papers in the scoping review was considered impractical because of the high number of papers. Moreover, only papers published in English were included. Finally, owing to the number of papers identified through the searches, only a random sample of 20% was screened by all coauthors.

Conclusions

This review offers a snapshot of evidence on what public involvement and engagement in big data research could look like. It is limited, as it was largely based on discussion papers, but useful, as evidence on how these involvement and engagement activities could be delivered and what type of outcomes they could produce was provided. The field would benefit from further research and evaluation of involvement and engagement activities in big data through primary research. Owing to the ongoing development of big data research, it is likely that these would need to be updated on a regular basis, but nevertheless, such research could provide further insights into how to meaningfully involve and engage the public in big data research.

Acknowledgments

PT was a PhD student supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration North West Coast (ARC NWC) and based at the University of Liverpool. SER is partly funded by the NIHR ARC NWC. This report is an independent research study funded by the ARC NWC. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors would like to thank Dr Kate Fleming for assisting at the data extraction stage.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. [[DOCX File , 27 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

Search strategy as published in the protocol paper. [[DOCX File , 20 KB-Multimedia Appendix 2](#)]

Multimedia Appendix 3

Data extraction form. [[DOCX File , 23 KB-Multimedia Appendix 3](#)]

References

1. 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. *Int J Popul Data Sci.* Mar 12, 2019;4(1):586-512. [[FREE Full text](#)] [doi: [10.23889/ijpds.v4i1.586](https://doi.org/10.23889/ijpds.v4i1.586)] [Medline: [34095528](https://pubmed.ncbi.nlm.nih.gov/34095528/)]
2. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: a systematic review. *Int J Med Inform.* Jun 2018;114:57-65. [doi: [10.1016/j.jmedinf.2018.03.013](https://doi.org/10.1016/j.jmedinf.2018.03.013)] [Medline: [29673604](https://pubmed.ncbi.nlm.nih.gov/29673604/)]
3. Newburn M, Scanlon M, Plachcinski R, Jill Macfarlane A. Involving service users in the Birth Timing project, a data linkage study analysing the timing of births and their outcomes. *Int J Popul Data Sci.* Nov 02, 2020;5(3):1366. [[FREE Full text](#)] [doi: [10.23889/ijpds.v5i3.1366](https://doi.org/10.23889/ijpds.v5i3.1366)] [Medline: [34007886](https://pubmed.ncbi.nlm.nih.gov/34007886/)]
4. Aitken M, Porteous C, Creamer E, Cunningham-Burley S. Who benefits and how? Public expectations of public benefits from data-intensive health research. *Big Data Soc.* Dec 06, 2018;5(2):205395171881672. [doi: [10.1177/2053951718816724](https://doi.org/10.1177/2053951718816724)]

5. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst.* 2014;2:3. [FREE Full text] [doi: [10.1186/2047-2501-2-3](https://doi.org/10.1186/2047-2501-2-3)] [Medline: [25825667](https://pubmed.ncbi.nlm.nih.gov/25825667/)]
6. Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. *Ethn Dis.* 2017;27(2):95-106. [FREE Full text] [doi: [10.18865/ed.27.2.95](https://doi.org/10.18865/ed.27.2.95)] [Medline: [28439179](https://pubmed.ncbi.nlm.nih.gov/28439179/)]
7. Lipworth W, Mason PH, Kerridge I, Ioannidis JP. Ethics and epistemology in big data research. *J Bioeth Inq.* Dec 20, 2017;14(4):489-500. [doi: [10.1007/s11673-017-9771-3](https://doi.org/10.1007/s11673-017-9771-3)] [Medline: [28321561](https://pubmed.ncbi.nlm.nih.gov/28321561/)]
8. Taylor M. information governance as a force for good? Lessons to be learnt from Care.data. *SCRIPTed.* Apr 2014;11(1):1-10. [doi: [10.2966/scrip.110114.1](https://doi.org/10.2966/scrip.110114.1)]
9. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. *J Med Ethics.* May 2015;41(5):404-409. [FREE Full text] [doi: [10.1136/medethics-2014-102374](https://doi.org/10.1136/medethics-2014-102374)] [Medline: [25617016](https://pubmed.ncbi.nlm.nih.gov/25617016/)]
10. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: a qualitative study. *J Med Internet Res.* Apr 15, 2016;18(4):e66. [FREE Full text] [doi: [10.2196/jmir.5011](https://doi.org/10.2196/jmir.5011)] [Medline: [27083521](https://pubmed.ncbi.nlm.nih.gov/27083521/)]
11. Shaw JA, Sethi N, Cassel CK. Social license for the use of big data in the COVID-19 era. *NPJ Digit Med.* Oct 02, 2020;3(1):128. [FREE Full text] [doi: [10.1038/s41746-020-00342-y](https://doi.org/10.1038/s41746-020-00342-y)] [Medline: [33083563](https://pubmed.ncbi.nlm.nih.gov/33083563/)]
12. Ford E, Boyd A, Bowles JK, Havard A, Aldridge RW, Curcin V, et al. Our data, our society, our health: a vision for inclusive and transparent health data science in the United Kingdom and beyond. *Learn Health Syst.* Jul 25, 2019;3(3):e10191. [FREE Full text] [doi: [10.1002/lrh2.10191](https://doi.org/10.1002/lrh2.10191)] [Medline: [31317072](https://pubmed.ncbi.nlm.nih.gov/31317072/)]
13. 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.* Feb 07, 2018;16(1):5. [FREE Full text] [doi: [10.1186/s12961-018-0282-4](https://doi.org/10.1186/s12961-018-0282-4)] [Medline: [29415734](https://pubmed.ncbi.nlm.nih.gov/29415734/)]
14. Muller SH, Kalkman S, van Thiel GJ, Mostert M, van Delden JJ. The social licence for data-intensive health research: towards co-creation, public value and trust. *BMC Med Ethics.* Aug 10, 2021;22(1):110. [FREE Full text] [doi: [10.1186/s12910-021-00677-5](https://doi.org/10.1186/s12910-021-00677-5)] [Medline: [34376204](https://pubmed.ncbi.nlm.nih.gov/34376204/)]
15. Dixon-Woods M, Ashcroft RE. Regulation and the social licence for medical research. *Med Health Care Philos.* Dec 17, 2008;11(4):381-391. [doi: [10.1007/s11019-008-9152-0](https://doi.org/10.1007/s11019-008-9152-0)] [Medline: [18633729](https://pubmed.ncbi.nlm.nih.gov/18633729/)]
16. Jones KH, Heys S, Thompson R, Cross L, Ford D. Public involvement and engagement in the work of a data safe haven: a case study of the SAIL databank. *Int J Popul Data Sci.* Aug 24, 2020;5(3):1371. [FREE Full text] [doi: [10.23889/ijpds.v5i3.1371](https://doi.org/10.23889/ijpds.v5i3.1371)] [Medline: [33644414](https://pubmed.ncbi.nlm.nih.gov/33644414/)]
17. Jones KH, Ford DV, Thompson S, Lyons RA. A profile of the SAIL databank on the UK secure research platform. *Int J Popul Data Sci.* Nov 20, 2019;4(2):1134. [FREE Full text] [doi: [10.23889/ijpds.v4i2.1134](https://doi.org/10.23889/ijpds.v4i2.1134)] [Medline: [34095541](https://pubmed.ncbi.nlm.nih.gov/34095541/)]
18. Jones KH, Mc Nerney CL, Ford DV. Involving consumers in the work of a data linkage research unit. *Int J Consumer Studies.* Oct 07, 2013;38(1):45-51. [FREE Full text] [doi: [10.1111/jics.12062](https://doi.org/10.1111/jics.12062)]
19. 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 Med Ethics.* Nov 10, 2016;17(1):73. [FREE Full text] [doi: [10.1186/s12910-016-0153-x](https://doi.org/10.1186/s12910-016-0153-x)] [Medline: [27832780](https://pubmed.ncbi.nlm.nih.gov/27832780/)]
20. Stockdale J, Cassell J, Ford E. "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 Res.* Jan 17, 2019;3:6. [doi: [10.12688/wellcomeopenres.13531.2](https://doi.org/10.12688/wellcomeopenres.13531.2)]
21. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *J Med Ethics.* Nov 12, 2019;48(1):3-13. [FREE Full text] [doi: [10.1136/medethics-2019-105651](https://doi.org/10.1136/medethics-2019-105651)] [Medline: [31719155](https://pubmed.ncbi.nlm.nih.gov/31719155/)]
22. Howe N, Giles E, Newbury-Birch D, McColl E. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *J Health Serv Res Policy.* Apr 13, 2018;23(2):123-133. [doi: [10.1177/1355819617751555](https://doi.org/10.1177/1355819617751555)] [Medline: [29653503](https://pubmed.ncbi.nlm.nih.gov/29653503/)]
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol.* Feb 2005;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
24. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. *J Clin Epidemiol.* Dec 2014;67(12):1291-1294. [doi: [10.1016/j.jclinepi.2014.03.013](https://doi.org/10.1016/j.jclinepi.2014.03.013)] [Medline: [25034198](https://pubmed.ncbi.nlm.nih.gov/25034198/)]
25. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci.* 2010;5:69. [FREE Full text] [doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)] [Medline: [20854677](https://pubmed.ncbi.nlm.nih.gov/20854677/)]
26. Teodorowski P, Rodgers S, 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 Expect.* Apr 2023;26(2):882-891. [FREE Full text] [doi: [10.1111/hex.13713](https://doi.org/10.1111/hex.13713)] [Medline: [36691930](https://pubmed.ncbi.nlm.nih.gov/36691930/)]
27. Teodorowski P, Jones E, Tahir N, Ahmed S, Frith L. 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.* Aug 19, 2021;11(8):e050167. [FREE Full text] [doi: [10.1136/bmjopen-2021-050167](https://doi.org/10.1136/bmjopen-2021-050167)] [Medline: [34413107](https://pubmed.ncbi.nlm.nih.gov/34413107/)]

28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* Oct 02, 2018;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
29. Mockford C, Staniszewska S, Griffiths F, Herron-Marx S. The impact of patient and public involvement on UK NHS health care: a systematic review. *Int J Qual Health Care.* Feb 2012;24(1):28-38. [FREE Full text] [doi: [10.1093/intqhc/mzr066](https://doi.org/10.1093/intqhc/mzr066)] [Medline: [22109631](https://pubmed.ncbi.nlm.nih.gov/22109631/)]
30. Islam S, Small N. An annotated and critical glossary of the terminology of inclusion in healthcare and health research. *Res Involv Engagem.* Apr 20, 2020;6(1):14. [FREE Full text] [doi: [10.1186/s40900-020-00186-6](https://doi.org/10.1186/s40900-020-00186-6)] [Medline: [32337067](https://pubmed.ncbi.nlm.nih.gov/32337067/)]
31. Dawson S, Campbell SM, Giles SJ, Morris RL, Cheraghi-Sohi S. Black and minority ethnic group involvement in health and social care research: a systematic review. *Health Expect.* Feb 15, 2018;21(1):3-22. [FREE Full text] [doi: [10.1111/hex.12597](https://doi.org/10.1111/hex.12597)] [Medline: [28812330](https://pubmed.ncbi.nlm.nih.gov/28812330/)]
32. Harrison JD, Auerbach AD, Anderson W, Fagan M, Carnie M, Hanson C, et al. Patient stakeholder engagement in research: a narrative review to describe foundational principles and best practice activities. *Health Expect.* Jun 13, 2019;22(3):307-316. [FREE Full text] [doi: [10.1111/hex.12873](https://doi.org/10.1111/hex.12873)] [Medline: [30761699](https://pubmed.ncbi.nlm.nih.gov/30761699/)]
33. Lalani M, Baines R, Bryce M, Marshall M, Mead S, Barasi S, et al. Patient and public involvement in medical performance processes: a systematic review. *Health Expect.* Apr 2019;22(2):149-161. [FREE Full text] [doi: [10.1111/hex.12852](https://doi.org/10.1111/hex.12852)] [Medline: [30548359](https://pubmed.ncbi.nlm.nih.gov/30548359/)]
34. Arnstein SR. A ladder of citizen participation. *J Am Inst Plann.* Jul 1969;35(4):216-224. [doi: [10.1080/01944366908977225](https://doi.org/10.1080/01944366908977225)]
35. What is public involvement in research? INVOLVE. 2020. URL: <https://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/> [accessed 2024-09-21]
36. Boote J, Baird W, Sutton A. Public involvement in the systematic review process in health and social care: a narrative review of case examples. *Health Policy.* Oct 2011;102(2-3):105-116. [doi: [10.1016/j.healthpol.2011.05.002](https://doi.org/10.1016/j.healthpol.2011.05.002)] [Medline: [21641075](https://pubmed.ncbi.nlm.nih.gov/21641075/)]
37. Boote J, Baird W, Sutton A. Involving the public in systematic reviews: a narrative review of organizational approaches and eight case examples. *J Comp Eff Res.* Sep 2012;1(5):409-420. [FREE Full text] [doi: [10.2217/ce.12.46](https://doi.org/10.2217/ce.12.46)] [Medline: [24236418](https://pubmed.ncbi.nlm.nih.gov/24236418/)]
38. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* Jul 21, 2009;6(7):e1000097. [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
39. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* Jan 01, 2015;4(1):1. [FREE Full text] [doi: [10.1186/2046-4053-4-1](https://doi.org/10.1186/2046-4053-4-1)] [Medline: [25554246](https://pubmed.ncbi.nlm.nih.gov/25554246/)]
40. Baart IL, Abma TA. Patient participation in fundamental psychiatric genomics research: a Dutch case study. *Health Expect.* Sep 2011;14(3):240-249. [FREE Full text] [doi: [10.1111/j.1369-7625.2010.00634.x](https://doi.org/10.1111/j.1369-7625.2010.00634.x)] [Medline: [20860783](https://pubmed.ncbi.nlm.nih.gov/20860783/)]
41. Ballantyne A, Style R. Health data research in New Zealand: updating the ethical governance framework. *N Z Med J.* Oct 27, 2017;130(1464):64-71. [Medline: [29073658](https://pubmed.ncbi.nlm.nih.gov/29073658/)]
42. Ballantyne A, Stewart C. Big data and public-private partnerships in healthcare and research: the application of an ethics framework for big data in health and research. *Asian Bioeth Rev.* Sep 30, 2019;11(3):315-326. [FREE Full text] [doi: [10.1007/s41649-019-00100-7](https://doi.org/10.1007/s41649-019-00100-7)] [Medline: [33717319](https://pubmed.ncbi.nlm.nih.gov/33717319/)]
43. Beyer KM, Comstock S, Seagren R. 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.* Dec 2010;35(6):635-644. [doi: [10.1007/s10900-010-9254-5](https://doi.org/10.1007/s10900-010-9254-5)] [Medline: [20352481](https://pubmed.ncbi.nlm.nih.gov/20352481/)]
44. Bharti N, O'Donovan C, Smallman M, Wilson J. Public trust, deliberative engagement and health data projects: beyond legal provisions. *Engag Sci Technol Soc.* Oct 05, 2021;7(1):125-133. [doi: [10.17351/ests2021.1197](https://doi.org/10.17351/ests2021.1197)]
45. Bot BM, Wilbanks JT, Mangravite LM. Assessing the consequences of decentralizing biomedical research. *Big Data Soc.* Jun 11, 2019;6(1):205395171985385. [doi: [10.1177/2053951719853858](https://doi.org/10.1177/2053951719853858)]
46. Coulter A. Patient trust in plans to share primary care data. *BMJ.* Jun 04, 2021;373:n1413. [doi: [10.1136/bmj.n1413](https://doi.org/10.1136/bmj.n1413)] [Medline: [34088678](https://pubmed.ncbi.nlm.nih.gov/34088678/)]
47. Dankar FK, Ptitsyn A, Dankar SK. The development of large-scale de-identified biomedical databases in the age of genomics-principles and challenges. *Hum Genomics.* Dec 10, 2018;12(1):19. [FREE Full text] [doi: [10.1186/s40246-018-0147-5](https://doi.org/10.1186/s40246-018-0147-5)] [Medline: [29636096](https://pubmed.ncbi.nlm.nih.gov/29636096/)]
48. de Freitas C, Amorim M, Machado H, Leão Teles E, Baptista MJ, Renedo A, et al. 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.* Mar 15, 2021;11(3):e044289. [FREE Full text] [doi: [10.1136/bmjopen-2020-044289](https://doi.org/10.1136/bmjopen-2020-044289)] [Medline: [33722870](https://pubmed.ncbi.nlm.nih.gov/33722870/)]
49. 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.* Mar 01, 2019;47(1):70-87. [FREE Full text] [doi: [10.1177/1073110519840486](https://doi.org/10.1177/1073110519840486)] [Medline: [30994071](https://pubmed.ncbi.nlm.nih.gov/30994071/)]

50. Duchange N, Darquy S, d'Audiffret D, Callies I, Lapointe AS, Loeve B, et al. Ethical management in the constitution of a European database for leukodystrophies rare diseases. *Eur J Paediatr Neurol*. Sep 2014;18(5):597-603. [FREE Full text] [doi: [10.1016/j.ejpn.2014.04.002](https://doi.org/10.1016/j.ejpn.2014.04.002)] [Medline: [24786336](https://pubmed.ncbi.nlm.nih.gov/24786336/)]
51. Erikainen S, Friesen P, Rand L, Jongsma K, Dunn M, Sorbie A, et al. Public involvement in the governance of population-level biomedical research: unresolved questions and future directions. *J Med Ethics*. Oct 06, 2020;47:522-525. [FREE Full text] [doi: [10.1136/medethics-2020-106530](https://doi.org/10.1136/medethics-2020-106530)] [Medline: [33023977](https://pubmed.ncbi.nlm.nih.gov/33023977/)]
52. Evans EA, Delorme E, Cyr K, Goldstein DM. A qualitative study of big data and the opioid epidemic: recommendations for data governance. *BMC Med Ethics*. Oct 21, 2020;21(1):101. [FREE Full text] [doi: [10.1186/s12910-020-00544-9](https://doi.org/10.1186/s12910-020-00544-9)] [Medline: [33087123](https://pubmed.ncbi.nlm.nih.gov/33087123/)]
53. Fernando B, King M, Sumathipala A. Advancing good governance in data sharing and biobanking - international aspects. *Wellcome Open Res*. Nov 22, 2019;4:184. [FREE Full text] [doi: [10.12688/wellcomeopenres.15540.1](https://doi.org/10.12688/wellcomeopenres.15540.1)] [Medline: [31950088](https://pubmed.ncbi.nlm.nih.gov/31950088/)]
54. Fleurence RL, Beal AC, Sheridan SE, Johnson LB, Selby JV. Patient-powered research networks aim to improve patient care and health research. *Health Aff (Millwood)*. Jul 2014;33(7):1212-1219. [doi: [10.1377/hlthaff.2014.0113](https://doi.org/10.1377/hlthaff.2014.0113)] [Medline: [25006148](https://pubmed.ncbi.nlm.nih.gov/25006148/)]
55. Funnell S, Tanuseputro P, Letendre A, Bearskin LB, Walker J. "Nothing about us, without us." How community-based participatory research methods were adapted in an indigenous end-of-life study using previously collected data. *Can J Aging*. Jun 20, 2020;39(2):145-155. [doi: [10.1017/S0714980819000291](https://doi.org/10.1017/S0714980819000291)] [Medline: [31746723](https://pubmed.ncbi.nlm.nih.gov/31746723/)]
56. Gallier S, Price G, Pandya H, McCarmack G, James C, Ruane B, et al. 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 Care Inform*. Apr 13, 2021;28(1):e100294. [FREE Full text] [doi: [10.1136/bmjhci-2020-100294](https://doi.org/10.1136/bmjhci-2020-100294)] [Medline: [33849921](https://pubmed.ncbi.nlm.nih.gov/33849921/)]
57. Goytia CN, Kastenbaum I, Shelley D, Horowitz CR, Kaushal R. A tale of 2 constituencies: exploring patient and clinician perspectives in the age of big data. *Med Care*. Oct 2018;56 Suppl 10 Suppl 1(10 Suppl 1):S64-S69. [FREE Full text] [doi: [10.1097/MLR.0000000000000786](https://doi.org/10.1097/MLR.0000000000000786)] [Medline: [30074954](https://pubmed.ncbi.nlm.nih.gov/30074954/)]
58. Henare KL, Parker KE, Wihongi H, Blenkinsop C, Jansen R, Reid P, et al. Mapping a route to indigenous engagement in cancer genomic research. *Lancet Oncol*. Jun 2019;20(6):e327-e335. [doi: [10.1016/S1470-2045\(19\)30307-9](https://doi.org/10.1016/S1470-2045(19)30307-9)] [Medline: [31162106](https://pubmed.ncbi.nlm.nih.gov/31162106/)]
59. Hudson M, Garrison NA, Sterling R, Caron NR, Fox K, Yracheta J, et al. Rights, interests and expectations: indigenous perspectives on unrestricted access to genomic data. *Nat Rev Genet*. Jun 06, 2020;21(6):377-384. [doi: [10.1038/s41576-020-0228-x](https://doi.org/10.1038/s41576-020-0228-x)] [Medline: [32251390](https://pubmed.ncbi.nlm.nih.gov/32251390/)]
60. Hurt L, Ashfield-Watt P, Townson J, Heslop L, Copeland L, Atkinson MD, et al. Cohort profile: HealthWise Wales, a research register and population health data platform with linkage to national health service data sets in Wales. *BMJ Open*. Dec 02, 2019;9(12):e031705. [FREE Full text] [doi: [10.1136/bmjopen-2019-031705](https://doi.org/10.1136/bmjopen-2019-031705)] [Medline: [31796481](https://pubmed.ncbi.nlm.nih.gov/31796481/)]
61. 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. *Res Involv Engagem*. 2019;5:20. [FREE Full text] [doi: [10.1186/s40900-019-0152-4](https://doi.org/10.1186/s40900-019-0152-4)] [Medline: [31205751](https://pubmed.ncbi.nlm.nih.gov/31205751/)]
62. Kalkman S, Mostert M, Gerlinger C, van Delden JJ, van Thiel GJ. Responsible data sharing in international health research: a systematic review of principles and norms. *BMC Med Ethics*. Mar 28, 2019;20(1):21. [FREE Full text] [doi: [10.1186/s12910-019-0359-9](https://doi.org/10.1186/s12910-019-0359-9)] [Medline: [30922290](https://pubmed.ncbi.nlm.nih.gov/30922290/)]
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. *Front Psychiatry*. Jun 10, 2021;12:643914. [FREE Full text] [doi: [10.3389/fpsy.2021.643914](https://doi.org/10.3389/fpsy.2021.643914)] [Medline: [34177644](https://pubmed.ncbi.nlm.nih.gov/34177644/)]
64. Luna Puerta L, Kendall W, Davies B, Day S, Ward H. The reported impact of public involvement in biobanks: a scoping review. *Health Expect*. Aug 06, 2020;23(4):759-788. [FREE Full text] [doi: [10.1111/hex.13067](https://doi.org/10.1111/hex.13067)] [Medline: [32378306](https://pubmed.ncbi.nlm.nih.gov/32378306/)]
65. Manrique de Lara A, Peláez-Ballestas I. Big data and data processing in rheumatology: bioethical perspectives. *Clin Rheumatol*. Apr 2020;39(4):1007-1014. [doi: [10.1007/s10067-020-04969-w](https://doi.org/10.1007/s10067-020-04969-w)] [Medline: [32062767](https://pubmed.ncbi.nlm.nih.gov/32062767/)]
66. Milne R, Sorbie A, Dixon-Woods M. What can data trusts for health research learn from participatory governance in biobanks? *J Med Ethics*. May 19, 2022;48(5):323-328. [FREE Full text] [doi: [10.1136/medethics-2020-107020](https://doi.org/10.1136/medethics-2020-107020)] [Medline: [33741681](https://pubmed.ncbi.nlm.nih.gov/33741681/)]
67. Milne R, Brayne C. We need to think about data governance for dementia research in a digital era. *Alzheimers Res Ther*. Jan 31, 2020;12(1):17. [FREE Full text] [doi: [10.1186/s13195-020-0584-y](https://doi.org/10.1186/s13195-020-0584-y)] [Medline: [32005135](https://pubmed.ncbi.nlm.nih.gov/32005135/)]
68. Mourby MJ, Doidge J, Jones KH, Aidinlis S, Smith H, Bell J, et al. Health data linkage for UK public interest research: key obstacles and solutions. *Int J Popul Data Sci*. Apr 02, 2019;4(1):1093. [FREE Full text] [doi: [10.23889/ijpds.v4i1.1093](https://doi.org/10.23889/ijpds.v4i1.1093)] [Medline: [32935027](https://pubmed.ncbi.nlm.nih.gov/32935027/)]
69. Murtagh MJ, Blell MT, Butters OW, Cowley L, Dove ES, Goodman A, et al. Better governance, better access: practising responsible data sharing in the METADAC governance infrastructure. *Hum Genomics*. Apr 26, 2018;12(1):24. [FREE Full text] [doi: [10.1186/s40246-018-0154-6](https://doi.org/10.1186/s40246-018-0154-6)] [Medline: [29695297](https://pubmed.ncbi.nlm.nih.gov/29695297/)]

70. Nelson E, Burns F. Impact through engagement: co-production of administrative data research and the approach of the administrative data research centre Northern Ireland. *Int J Popul Data Sci.* Nov 10, 2020;5(3):1369. [FREE Full text] [doi: [10.23889/ijpds.v5i3.1369](https://doi.org/10.23889/ijpds.v5i3.1369)] [Medline: [34007887](https://pubmed.ncbi.nlm.nih.gov/34007887/)]
71. Nunn JS, Gwynne K, Gray S, Lacaze P. Involving people affected by a rare condition in shaping future genomic research. *Res Involv Engagem.* Mar 15, 2021;7(1):14. [FREE Full text] [doi: [10.1186/s40900-021-00256-3](https://doi.org/10.1186/s40900-021-00256-3)] [Medline: [33722276](https://pubmed.ncbi.nlm.nih.gov/33722276/)]
72. O'Doherty KC, Burgess MM, Edwards K, Gallagher RP, Hawkins AK, Kaye J, et al. From consent to institutions: designing adaptive governance for genomic biobanks. *Soc Sci Med.* Aug 2011;73(3):367-374. [doi: [10.1016/j.socscimed.2011.05.046](https://doi.org/10.1016/j.socscimed.2011.05.046)] [Medline: [21726926](https://pubmed.ncbi.nlm.nih.gov/21726926/)]
73. O'Doherty KC, Shabani M, Dove ES, Bentzen HB, Borry P, Burgess MM, et al. Toward better governance of human genomic data. *Nat Genet.* Jan 07, 2021;53(1):2-8. [FREE Full text] [doi: [10.1038/s41588-020-00742-6](https://doi.org/10.1038/s41588-020-00742-6)] [Medline: [33414545](https://pubmed.ncbi.nlm.nih.gov/33414545/)]
74. Ohno-Machado L, Agha Z, Bell DS, Dahm L, Day ME, Doctor JN, et al. pSCANNER: patient-centered scalable national network for effectiveness research. *J Am Med Inform Assoc.* Jul 01, 2014;21(4):621-626. [FREE Full text] [doi: [10.1136/amiajnl-2014-002751](https://doi.org/10.1136/amiajnl-2014-002751)] [Medline: [24780722](https://pubmed.ncbi.nlm.nih.gov/24780722/)]
75. Omar MI, Roobol MJ, Ribal MJ, Abbott T, Agapow PM, Araujo S, et al. Introducing PIONEER: a project to harness big data in prostate cancer research. *Nat Rev Urol.* Jun 2020;17(6):351-362. [doi: [10.1038/s41585-020-0324-x](https://doi.org/10.1038/s41585-020-0324-x)] [Medline: [32461687](https://pubmed.ncbi.nlm.nih.gov/32461687/)]
76. Paprica PA, Sutherland E, Smith A, Brudno M, Cartagena RG, Crichlow M, et al. Essential requirements for establishing and operating data trusts: practical guidance co-developed by representatives from fifteen canadian organizations and initiatives. *Int J Popul Data Sci.* Aug 24, 2020;5(1):1353. [FREE Full text] [doi: [10.23889/ijpds.v5i1.1353](https://doi.org/10.23889/ijpds.v5i1.1353)] [Medline: [33644412](https://pubmed.ncbi.nlm.nih.gov/33644412/)]
77. Patel R, Irving J, Brinn A, Broadbent M, Shetty H, Pritchard M, et al. Impact of the COVID-19 pandemic on remote mental healthcare and prescribing in psychiatry: an electronic health record study. *BMJ Open.* Mar 30, 2021;11(3):e046365-e046363. [FREE Full text] [doi: [10.1136/bmjopen-2020-046365](https://doi.org/10.1136/bmjopen-2020-046365)] [Medline: [33785494](https://pubmed.ncbi.nlm.nih.gov/33785494/)]
78. Pavlenko E, Strech D, Langhof H. Implementation of data access and use procedures in clinical data warehouses. A systematic review of literature and publicly available policies. *BMC Med Inform Decis Mak.* Jul 11, 2020;20(1):157. [FREE Full text] [doi: [10.1186/s12911-020-01177-z](https://doi.org/10.1186/s12911-020-01177-z)] [Medline: [32652989](https://pubmed.ncbi.nlm.nih.gov/32652989/)]
79. Rowe R, Carroll SR, Healy C, Rodriguez-Lonebear D, Walker JD. The SEEDS of indigenous population health data linkage. *Int J Popul Data Sci.* Jun 22, 2021;6(1):1417. [FREE Full text] [doi: [10.23889/ijpds.v6i1.1417](https://doi.org/10.23889/ijpds.v6i1.1417)] [Medline: [34212119](https://pubmed.ncbi.nlm.nih.gov/34212119/)]
80. Sleigh J, Vayena E. Public engagement with health data governance: the role of visuality. *Humanit Soc Sci Commun.* Jun 18, 2021;8(1):149. [doi: [10.1057/S41599-021-00826-6](https://doi.org/10.1057/S41599-021-00826-6)]
81. 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. *Int J Popul Data Sci.* May 07, 2019;4(1):1103. [FREE Full text] [doi: [10.23889/ijpds.v4i1.1103](https://doi.org/10.23889/ijpds.v4i1.1103)] [Medline: [34095532](https://pubmed.ncbi.nlm.nih.gov/34095532/)]
82. Tindana P, de Vries J, Campbell M, Littler K, Seeley J, Marshall P, et al. Community engagement strategies for genomic studies in Africa: a review of the literature. *BMC Med Ethics.* Apr 12, 2015;16:24. [FREE Full text] [doi: [10.1186/s12910-015-0014-z](https://doi.org/10.1186/s12910-015-0014-z)] [Medline: [25889051](https://pubmed.ncbi.nlm.nih.gov/25889051/)]
83. 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. *Int J Popul Data Sci.* Apr 16, 2020;5(3):1356. [FREE Full text] [doi: [10.23889/ijpds.v5i3.1356](https://doi.org/10.23889/ijpds.v5i3.1356)] [Medline: [34007884](https://pubmed.ncbi.nlm.nih.gov/34007884/)]
84. Vayena E, Blasimme A. Biomedical big data: new models of control over access, use and governance. *J Bioeth Inq.* Dec 5, 2017;14(4):501-513. [FREE Full text] [doi: [10.1007/s11673-017-9809-6](https://doi.org/10.1007/s11673-017-9809-6)] [Medline: [28983835](https://pubmed.ncbi.nlm.nih.gov/28983835/)]
85. Weich S, Duncan C, Bhui K, Canaway A, Crepez-Keay D, Keown P, et al. 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.* Oct 18, 2018;8(10):e024193. [FREE Full text] [doi: [10.1136/bmjopen-2018-024193](https://doi.org/10.1136/bmjopen-2018-024193)] [Medline: [30341141](https://pubmed.ncbi.nlm.nih.gov/30341141/)]
86. Willison DJ, Trowbridge J, Greiver M, Keshavjee K, Mumford D, Sullivan F. Participatory governance over research in an academic research network: the case of Diabetes Action Canada. *BMJ Open.* Apr 20, 2019;9(4):e026828. [FREE Full text] [doi: [10.1136/bmjopen-2018-026828](https://doi.org/10.1136/bmjopen-2018-026828)] [Medline: [31005936](https://pubmed.ncbi.nlm.nih.gov/31005936/)]
87. Xafis V, Labude MK. Openness in big data and data repositories: the application of an ethics framework for big data in health and research. *Asian Bioeth Rev.* Sep 01, 2019;11(3):255-273. [FREE Full text] [doi: [10.1007/s41649-019-00097-z](https://doi.org/10.1007/s41649-019-00097-z)] [Medline: [33717315](https://pubmed.ncbi.nlm.nih.gov/33717315/)]
88. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. *BMJ.* Aug 02, 2017;358:j3453. [FREE Full text] [doi: [10.1136/bmj.j3453](https://doi.org/10.1136/bmj.j3453)] [Medline: [28768629](https://pubmed.ncbi.nlm.nih.gov/28768629/)]
89. PEDRI: public involvement and engagement best practice draft standards for the use of data for research and statistics. ADR UK & Economic and Social Research Council. 2023. URL: https://www.adruk.org/fileadmin/uploads/adruk/Documents/PE_reports_and_documents/PEDRI-Best-Practice-Standards.pdf [accessed 2024-04-10]
90. NIHR. National Standards for Public InvolvementNIHR announces new standards for public involvement in research. National Institutes for Health and Care Research (NIHR). 2019. URL: <https://www.nihr.ac.uk/news/nihr-announces-new-standards-for-public-involvement-in-research/23830> [accessed 2024-04-29]

91. Pii KH, Schou LH, Piil K, Jarden M. Current trends in patient and public involvement in cancer research: a systematic review. *Health Expect*. Mar 2019;22(1):3-20. [FREE Full text] [doi: [10.1111/hex.12841](https://doi.org/10.1111/hex.12841)] [Medline: [30378234](https://pubmed.ncbi.nlm.nih.gov/30378234/)]
92. Bailey WB, Twins B, Wilkinson-Salamea C, Raidos D, Imafidon K, McGarry N. 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. 2021. URL: https://understandingpatientdata.org.uk/sites/default/files/2022-04/Diverse%20voices%20on%20Data%20-%20Main%20report_0.pdf [accessed 2024-04-29]
93. 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*. Oct 2014;17(5):637-650. [doi: [10.1111/j.1369-7625.2012.00795.x](https://doi.org/10.1111/j.1369-7625.2012.00795.x)] [Medline: [22809132](https://pubmed.ncbi.nlm.nih.gov/22809132/)]

Abbreviations

GRIPP2: Guidance for Reporting Involvement of Patients and the Public

NIHR: National Institute for Health and Care Research

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by M Hudson, S Woods; submitted 23.01.24; peer-reviewed by N Natafqi, A Paprica, M McCoy; comments to author 24.03.24; revised version received 06.05.24; accepted 22.06.24; published 16.08.24

Please cite as:

Teodorowski P, Jones E, Tahir N, Ahmed S, Rodgers SE, Frith L

Public Involvement and Engagement in Big Data Research: Scoping Review

J Particip Med 2024;16:e56673

URL: <https://jopm.jmir.org/2024/1/e56673>

doi: [10.2196/56673](https://doi.org/10.2196/56673)

PMID:

©Piotr Teodorowski, Elisa Jones, Naheed Tahir, Saiqa Ahmed, Sarah E Rodgers, Lucy Frith. Originally published in *Journal of Participatory Medicine* (<https://jopm.jmir.org>), 16.08.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *Journal of Participatory Medicine*, is properly cited. The complete bibliographic information, a link to the original publication on <https://jopm.jmir.org>, as well as this copyright and license information must be included.