

BMJ Open 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

Cláudia de Freitas ^{1,2}, Mariana Amorim,¹ Helena Machado,³ Elisa Leão Teles,⁴ Maria João Baptista,^{5,6} Alicia Renedo,⁷ Veerle Provoost,⁸ Susana Silva^{1,9}

To cite: de Freitas C, Amorim M, Machado H, *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* 2021;**11**:e044289. doi:10.1136/bmjopen-2020-044289

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-044289>).

Received 31 August 2020
Revised 17 February 2021
Accepted 26 February 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Cláudia de Freitas;
claudia.defreitas@ispup.up.pt

ABSTRACT

Introduction International policy imperatives for the public and patient involvement in the governance of health data coexist with conflicting cross-border policies on data sharing. This can challenge the planning and implementation of participatory data governance in healthcare services locally. Engaging with local stakeholders and understanding how their needs, values and preferences for governing health data can be articulated with policies made at the supranational level is crucial. This paper describes a protocol for a project that aims to coproduce a people-centred model for involving patients and the public in decision-making processes about the use and sharing of health data for rare diseases care and research.

Methods and analysis This multidisciplinary project draws on an explanatory sequential mixed-methods study. A hospital-based survey with patients, informal carers, health professionals and technical staff recruited at two reference centres for rare diseases in Portugal will be conducted first. The qualitative study will follow consisting of semi-structured interviews and scenario-based workshops with a subsample of the participant groups recruited at baseline. Quantitative data will be analysed using descriptive and inferential statistics. Inductive and deductive approaches will be combined to analyse the qualitative interviews. Data from scenario-based workshops will be iteratively compared using the constant comparison method to identify cross-cutting themes and categories.

Ethics and dissemination The Ethics Committee for Health from the University Hospital Centre São João/ Faculty of Medicine of University of Porto approved the study protocol (Ref. 99/19). Research findings will be disseminated at academic conferences and science promotion events, and through public meetings involving patient representatives, practitioners, policy-makers and students, a project website and peer-reviewed journal publications.

INTRODUCTION

How to govern health data has become a question of unprecedented relevance in

Strengths and limitations of this study

- This study will contribute to help reducing health data policy implementation gaps by coproducing a model for public and patient involvement in data governance in the context of rare diseases.
- Another of its strengths is its people centred, multi-disciplinary approach designed to elicit stakeholders' expertise to identify challenges to and devise strategies for enacting participatory health data governance.
- The project combines the strengths of quantitative and qualitative methods, maximising the latter's potential to facilitate the engagement of patients and the public with the topic of health data use and sharing.
- Although project funds only allow for data collection at two reference centres for rare diseases at one academic hospital centre, this study site was purposefully selected because it would enable access to stakeholders involved in national and international data sharing for research.
- Challenges associated with the study's multi-phase design include participant attrition between questionnaires, semistructured interviews and the scenario-based workshop, and lower response rates, namely by people who may be less resourced to participate or interested in research.

the golden age of data sharing.¹ Large-scale health data repositories aggregating multiple datasets hold enormous predictive and transformation potential to improve population health and well-being.² This is especially promising in the field of rare diseases, where transnational data sharing is likely to intensify as a result of increased sponsorship by the European Commission (EC).³ Yet how health data are shared and used also carries risks and causes public concern.⁴ International policy agencies recommend public participation in

data governance as a means to improve safeguards, reduce risks and increase the benefits arising from health data processing.^{5 6} But such recommendations exist alongside conflicting cross-border policies on data sharing^{7 8} that can thwart its implementation. Engaging with local stakeholders and understanding how their needs, values and preferences can be articulated with policies made at the supranational level is therefore a pressing concern.

Processing of large sets of health data can help to prevent, diagnose and cure hundreds of diseases.^{9 10} It can also improve clinical practice and quality of care, reduce public health threats and advance health policy goals.^{6 9 11 12} However, information security and privacy have been challenged by data breaches and mismanagement that can lead to participants' unduly identification.^{13 14} Unforeseen access to health data by private companies also raises concerns with confidentiality and inappropriate data use, including results commercialisation and undisclosed surveillance.^{4 15 16} These risks can erode public trust in the organisations that handle their data, reducing potential for innovation.^{4 17} To address these issues, several international legal instruments and recommendations have been produced.^{5 6 18} They emphasise the importance of involving the public in the development of a shared policy framework to govern the protection of privacy, data security and transnational data flows.^{5 6}

Rare diseases offer a paradigmatic example of the opportunities and challenges involved in the governance of health data. Their rareness, diversity and complexity demand collaboration and sharing of patient data between experts internationally. Data-sharing partnerships are key to overcome difficulties of conducting research in this field: case numbers are low and there is insufficient availability of technological tools needed to conduct genetic testing and genome sequencing.^{19 20} Yet this type of collaboration is difficult as data on rare diseases is sparse, dispersed and unsystematically collected and stored.²¹ To tackle this problem, the EC has recently funded the European Reference Networks (ERN) to set up rare disease patient registries.³ ERNs are expected to develop interoperable registries and to build synergetic partnerships for knowledge and data exchange.²² Realising these goals will prompt the flow of large amounts of data and technologies between member states.

Cross-border health data sharing raises additional ethical, legal and social challenges. These include unauthorised data reuse and unwanted return of incidental findings from research,^{4 23} which compromise individuals' autonomy and entitlement to an open future (eg, anticipating the diagnosis of an untreatable adult-onset disorder through return of unsolicited genetic testing results).²⁴ The EU General Data Protection Regulation, enforced since May 2018, aims to address some of these challenges, not least through the rights to be forgotten and to data portability.¹⁸ It does so on the footprints of two sets of policies on research and innovation. One of those policies promotes data sharing and harmonisation of

international rules toward a paradigm of 'open science'.²⁵ The other one is underlined by ethical concerns with autonomy and privacy protection and asserts the need to control data flows.²⁶ These data-sharing policies host mutually incompatible rules that can give rise to tensions when implemented^{7 27}: healthcare providers, for example, may struggle to combine the protection of patients' right to privacy while seeking to abide by the principles of open-access.

Participatory health data governance

Increasing support for participatory health data governance, whereby patients and other members of the public are involved in consultations and decision making about health data, is premised on the idea that this process can increase accountability in data processing, reduce unnecessary harms and enhance the fair distribution of the benefits of data use.^{5 6 8 28} Participatory data governance is also encouraged as a way of promoting dialogue about ethical, legal and social issues and imbuing research policy with socioethical sensitivity.²⁹⁻³² Collaboration between lay and research experts can infuse research with social values and facilitate the codesign of innovative solutions for complex problems.^{20 31} However, those collaborative relationships are not always easy to develop.³³ Careful attention is needed to create participatory spaces that are inclusive and harness everyone's potential for contribution.³⁴⁻³⁶ Competing interests, and an unbalanced distribution of the resources needed for meaningful participation, can cause disadvantaged groups to be excluded from decision making, which in turn risks to reproduce ethnic, age, gender and socioeconomic inequalities or to even generate new ones (eg, digital exclusion).^{34 37-40} There is thus an unequivocal need to assess the potential benefits and risks of public and patient involvement in health data governance and to develop a people-centred, bottom-up approach to participation that responds to local stakeholders' interests and needs. This project aims to reach this goal by collaborating with stakeholders engaged in care and research for rare diseases at an academic hospital centre in Portugal. Its specific aims are to:

1. Assess the needs and preferences of rare diseases patients, their informal carers, health professionals and technical staff concerning decision making about the use and sharing of health data for rare diseases care and research.
2. Understand those stakeholders' expectations of and perspectives about public and patient involvement in health data decision-making processes, including their views on the ethical, legal and social implications of participation.
3. Coproduce a people-centred model for public and patient involvement in health data governance.

DATAGov is led by a multidisciplinary team working at the crossroads of sociology, public health and clinical medicine and with long-term experience in collaborating with rare diseases community organisations and networks. The project will contribute to advance international

policy imperatives for public and patient involvement in health data governance and to improve its practice. It will do so by creating opportunities for stakeholders to discuss the implementation of participatory data governance in a healthcare setting and by eliciting their collective expertise to devise strategies to address emerging ethical, legal and social challenges.

METHODS AND ANALYSIS

Study design and setting

This project adopts an explanatory sequential mixed-methods design, collecting and analysing quantitative and qualitative data.⁴¹ The quantitative phase is first in the sequence and draws on a hospital-based survey with rare disease patients, informal carers, health professionals and technical staff recruited at the University Hospital Centre São João (UHCSJ), in Porto, Portugal. The qualitative phase relies on semistructured interviews and scenario-based workshops with a subsample of the participants recruited at baseline. Survey and interview data will be collected to examine stakeholders' needs, preferences, expectations and perspectives of public and patient involvement in health data decision making. These findings will be subsequently integrated to develop a set of scenarios for participatory health data governance, which will be presented at the scenario-based workshops. Representatives of all stakeholder groups will be invited to discuss the scenarios and to select a preferred model for involving patients and other members of the public in the governance of health data (see [table 1](#)). Integration of quantitative and qualitative findings will provide an understanding of the ethical, legal and social implications of participatory data governance.

Data will be collected in two reference centres for rare diseases at the UHCSJ. These research settings were purposefully sampled because: (1) they were among the first reference centres to be created in Portugal in 2016 and are acknowledged for their expertise and clinical practice; (2) they are in charge of overseeing patients

from the entire Northern Health Region of Portugal, which allows for larger study samples; (3) one is affiliated with an ERN and the other is awaiting a final decision on its application for membership and (4) they are involved in national and international research projects, which presents an excellent case to explore first-hand experiences of decision making regarding patient data use and sharing.

Study procedures and analysis

Hospital-based survey

Rare diseases patients, informal carers, health professionals and technical staff will be invited to participate in a survey conducted at the UHCSJ to assess their preferences and needs regarding decision making concerned with health data use and sharing.

During a 6-month period, patients aged 12 and above who attend a consultation in the reference centres at the UHCSJ, who can read and write in the Portuguese language and without cognitive impairment will be consecutively recruited to participate in the study, together with their informal carers. Medicine and nursing unit staff of the reference centres outpatient clinics will distribute study information leaflets to potential participants, or their guardians, after patients' biometric screening appointments or medical consultations. Two researchers will be at the hospital site to provide additional information and to invite patients and informal carers to independently fill in a self-report structured questionnaire, after obtaining informed consent. Underage participants will be invited to complete the questionnaire individually, after giving verbal consent and their carers signing the written consent form. Adult survey participants will be asked for consent to be contacted and invited to a semi-structured interview approximately 4 months later.

Directors and managers of relevant units and departments at the UHCSJ will be provided with recruitment emails, along with study information sheets, to invite health professionals from the reference centres' multidisciplinary teams (ie, doctors, nurses and allied health

Table 1 Mixed-methods research objectives, procedures and analysis

Objective	Study procedure	Sampling and participants	Data analysis
To assess stakeholders' needs and preferences concerning decision-making about health data use and sharing	Hospital-based survey	Consecutive sampling at two reference centres for rare diseases of an estimated 200 patients, 500 informal carers, 70 health professionals and 30 technical staff	Descriptive and regression analysis, adjusting for confounders
To understand stakeholders' expectations and perspectives about public and patient involvement in health data decision-making processes	Qualitative interviews	Subsample of quantitative study selected through heterogeneity sampling including an estimated 30 patients, 30 informal carers, 15 health professionals and 10 technical staff	Inductive and deductive analysis informed by the Modified Participation Chain Model
To coproduce a model for public and patient involvement in health data governance	Scenario-based workshops	Subsample of quantitative study selected through heterogeneity sampling including 8–12 participants per stakeholder group in the intragroup round and three participants per stakeholder group in the mixed-group round (total of 12 participants)	Deductive and inductive approach to scenario development followed by inductive analysis of workshops data



professionals) and technical staff to partake in the survey. Questionnaires will be delivered to potential participants after obtaining informed consent by two researchers over a 2-month period.

Questionnaire development was informed by current literature and a review of existing questionnaires related to the research topic. Professional experts from the health and social sciences (eg, general practitioner, psychologist, specialist doctor, researcher) were asked to pretest the structured questionnaire, which was revised and subsequently piloted with a group of patients (n=5) and informal carers (n=22). This resulted in modifications to wording and in some items being removed. The final questionnaire for patients and informal carers includes 38 questions covering four major topics: (1) positioning regarding data use and sharing for care and research purposes; (2) preferences and needs regarding data use and sharing; (3) opinions about public and patient involvement in health data governance and (4) sociodemographic characteristics. The questionnaire was adapted to accommodate professionals and technical staff, resulting in a slightly shorter version for these groups. A translation of the full questionnaire by the authors is available as online supplemental material. Questionnaires require an average of 15 min to fill in. We expect participation by 200 patients, 500 informal carers and 70 health professionals and 30 technical staff.

Statistical analyses will be performed using IBM SPSS V.25.0. Data will be described as counts and proportions for categorical variables, mean and SD for normally distributed continuous variables, and median and IQR for non-normally distributed continuous variables. Different analytic approaches will be performed to tackle the specific objectives established. Means will be compared with the Student's t-test or Mann-Whitney test, as appropriate, and differences in frequencies and proportions of categorical variables will be assessed using the χ^2 test or the Fisher's exact test. The OR and respective 95% CI will be calculated by logistic regression models to estimate the associations between explanatory variables (eg, sex, age, educational level, working status, occupation, perceived income adequacy, involvement with patient associations, satisfaction with own health, social trust) and the outcomes, which will include willingness to share health data for care and research purposes, importance given to participation in decision making about data use and sharing, preferred forms of consent to data use and willingness for and preferred modes of public and patient involvement in health data governance in a healthcare setting.

Qualitative interviews

To deepen our understanding of stakeholders' expectations and perspectives of public and patient involvement in health data decision making, semistructured interviews will be conducted with a subsample of the adult participants who completed the questionnaire. Approximately 4 months after participation in the survey, patients and

informal carers who consented to being contacted will be invited by telephone or email to take part in an individual qualitative interview. Health professionals and technical staff will be invited to participate through email. Heterogeneity sampling will be employed to obtain maximum variation in perspectives and experiences and allow theoretical saturation to be reached.^{42 43} Interviews will be held at a time and place of participants' preference (eg, university department, work place, home). Participant information sheets and consent forms will be provided alongside with a request to audiorecord the interviews.

Questions for the interview guide are exploratory and informed by existing literature. They address four main topics: (1) personal experience with involvement in patient organisations and/or illness; (2) facilitators of and barriers to health data sharing; (3) views about individual participation in decision-making concerning health data use and sharing and, (4) perspectives about public and patient involvement in health data governance. Interviews are expected to last between 30 min and 1 hour. We estimate to conduct interviews with 30 patients, 30 informal carers, 15 health professionals and 10 technical staff.

Data analysis will proceed hand in hand with data collection for timely identification of informational redundancy and subsequent discontinuation of sampling.^{42 44} Interview transcripts will be analysed iteratively by two researchers with the support of NVivo V.12. Inductive analysis will be performed first using open, selective and axial coding to facilitate the emergence of categories and core themes.⁴³ Deductive analysis will follow drawing on relevant theoretical frameworks, namely the Modified Participation Chain Model (MPCM).³⁴ The MPCM asserts that public involvement in health decision making is influenced by a set of mutually constitutive factors that evolve through participation when the right institutional dynamics are in place. These factors include individual and collective motivations, mobilisation efforts, resources and the dynamics of participation. They influence not only people's willingness to get involved but also their capacity and resolve to participate. These acumens will be used to load inductive themes with theoretical sensitivity.⁴³ Triangulation of interview data sources and of data analysts will be employed to ensure the rigour and quality of the findings.⁴⁵

Stakeholders' views will provide in depth insight in:

- ▶ How they may be driven to or detracted from data use and sharing for care and research and whether that is explained by perceived benefits and risks?
- ▶ Their stand concerning the involvement of different publics in collective decision-making about health data.
- ▶ Their views about the ethical, legal and social implications of participatory health data governance and proposed solutions to overcome potential challenges.

Scenario-based workshops

Scenario-based workshops⁴⁶ will be held to coproduce a people-centred model for public and patient involvement

in health data governance. This foresight method was chosen for its potential to facilitate the engagement of patients and the public in the development of health policy at local and national levels^{46–49} and it will be used to incorporate participants' collective expertise into a model for participatory data governance.

Survey and interview data will be integrated and combined with extant literature to inform the design of a set of scenarios for involving different stakeholders in collective decision-making processes concerned with health data. The scenarios will be presented in workshops with representatives from rare diseases patients, informal carers, health professionals and technical staff to enable critical dialogue. The scenario-based workshops will be carried out in two rounds: (1) round 1 will entail four intra-group workshops conducted with each stakeholder group separately (8–12 participants). Participants will be asked to evaluate the relevance and feasibility of each scenario and to assess its potential ethical, legal and social implications; (2) round 2 will entail two mixed-group workshops including a similar number of representatives from each stakeholder group (n=3) in a total of 12 participants. Stakeholders' feedback from round 1 will be used to make revisions to the first set of scenarios. The revised scenarios will be offered for discussion in round 2, in which participants will be asked to select a preferred model for public and patient involvement in health data governance.

Round 1 participants will be selected per stakeholder group and from the baseline sample through heterogeneity sampling, stratified by gender and educational level. They will be invited to participate by email or telephone, following their expressed preferences. A subsample of round 1 participants will be randomly selected to take part in round 2. The scenario-based workshops will be held in the venues considered most accessible by participants (eg, university department). They will be moderated by an experienced researcher supported by two observers who will carry out observation and note-taking. They are expected to last approximately 2 hours. Permission to audiorecord the workshops will be requested. Recordings will be transcribed verbatim and used for interpretational analysis with the assistance of NVivo V.12. Following open, axial and selective coding,⁴³ data from each scenario-based workshop will be iteratively compared with data from other, same-round scenario-based workshops using the constant-comparison method⁵⁰ to identify cross-cutting core themes and categories.

The scenario-based workshops will offer insights about:

- ▶ What motivates stakeholders to get involved in collective decision making about health data?
- ▶ Which resources are needed to promote meaningful participation by all stakeholders involved?
- ▶ How stakeholders can be best informed about and mobilised to undertake opportunities for involvement in health data governance?
- ▶ How collaborative participation dynamics may be achieved so as to accommodate stakeholders' needs,

values and preferences and inform a people-centred model for participatory data governance?

Public and patient involvement

The project's research question was informed by a previous empirical study conducted by some of the authors in which rare diseases community experts expressed the need to play an active role in decisions about how their health data is used.²⁰ Representatives of patients, informal carers, health professionals and researchers were involved in questionnaire and interview guide validation through pretesting and piloting. Their feedback was used to adapt its contents and to keep their length to a feasible minimum. In addition, the following strategies will be used to promote inclusivity in study participation: participants will be recruited by health professionals in a consecutive way, after their medical appointments; research team members will support participants during questionnaire completion, clarifying any arising doubts and interviews and scenario-based workshops will accommodate participants' scheduling and place preferences (eg, weekends; close to work).

The project's ultimate aim is to facilitate the co-production of a model for participatory health data governance. This will be achieved through involvement of a variety of stakeholders in critical dialogue about how to govern health data using scenario-based workshops. Participatory exercises that engage stakeholders with different backgrounds and interests offer an opportunity to expand knowledge and effect change that matters to increasingly diverse constituencies.³⁴ However, they may also cause the powerful-powerless gap to become more salient if the voices of groups traditionally lacking in equitable social participation are silenced.^{38,51} To prevent the latter, patients and informal carers will be provided with a training in preparation for the scenario-based workshops. All participants will be informed of the purposes and dynamics (eg, moderators' and participants' roles; appropriate rules of conduct) of the scenario-based workshops prior to their start. Furthermore, intragroup scenario-based workshops will be held first, to ensure that the perspectives of all stakeholder groups are elicited and incorporated in the scenarios presented at the mixed-group workshops. All workshops will be facilitated by academic leads with experience in involving patients and the public in healthcare research.²⁰

Stakeholder representatives will also be involved in research findings dissemination through participation in public meetings and science promotion events. The project website will provide a contact platform that stakeholders and the general public can use to address questions and comments to the research team.

ETHICS AND DISSEMINATION

The study protocol received ethical approval from the Ethics Committee for Health from the UHCSJ/Faculty of Medicine of University of Porto (Ref. 99/19).



To ensure data confidentiality and protection, participants' identifying information will only be available to four research team members, who will sign a non-disclosure agreement. Other researchers will be provided pseudoanonymised data.⁵² Questionnaires, interviews and scenario-based workshops will be conducted in a private place after obtaining consent. Personal contacts, consent forms, questionnaires, audiorecordings and transcripts will be coded and kept separately in computer files protected by password and in locked cabinets. Audio and transcribed data will be securely stored for a period of 5 years and eliminated afterwards.

Survey and interview research findings will be reported first to stakeholders at the in-group scenario-based workshops to identify areas of convergence and tension and prepare for inter-group dialogue about participatory data governance. In addition, the findings will be disseminated to a broad audience through peer-reviewed journal publications, academic conferences, a project website and science promotion events involving patient representatives, practitioners, policy-makers and students.

The project was initially planned to last 36 months, starting in July 2018. However, an extension of 12 months was requested to compensate for fieldwork constraints caused by the COVID-19 pandemic. The project is thus expected to run from July 2018 to July 2022. Adjustments to data collection and dissemination, including the use of online tools (eg, online questionnaires, webinars), will be made if the measures undertaken locally to address the pandemic so require.

EXPECTED OUTCOMES AND RELEVANCE

International policy recommendations for public involvement in data governance⁶ exist amidst conflicting data-sharing policies.^{7,27} This can challenge the effective planning and implementation of participatory data governance in local healthcare services. DATAGov is set to coproduce a people-centred model for collective participation in decision-making about data use and sharing for rare diseases care and research. To that end, it will help build collective expertise by a wide range of stakeholders, including patients and lay members of the public, both to unravel the social, legal and ethical implications of participatory data governance and to devise bottom-up strategies to address emerging problems.

Project results will contribute to inform and improve the practice of public and patient involvement in health data governance, nationally and internationally, by expanding the evidence base on stakeholders' stands, concerns and preferred roles regarding data decision-making processes. Undeniably, DATAGov is a timely and much needed response to calls for moving away from 'hypothetical person-centred approaches that presume a universal 'reasonable person' standard' into empirically grounded accounts of data governance⁸ that acknowledge the heterogeneity of interests underlining the constitution of different publics.^{40,53} As such, it will help to reduce policy implementation gaps by facilitating the coproduction of a data governance model more attuned with citizens' expectations, needs and preferences.

Author affiliations

- ¹EPIUnit – Instituto de Saúde Pública, Universidade do Porto, Porto, Portugal
- ²Centre for Research and Studies in Sociology, University Institute of Lisbon (ISCTE-IUL), Lisboa, Portugal
- ³Communication and Society Research Centre (CECS), Institute of Social Sciences, University of Minho, Braga, Portugal
- ⁴Centro de Referência de Doenças Hereditárias do Metabolismo do Centro Hospitalar Universitário São João, Porto, Portugal
- ⁵Centro de Referência de Cardiopatias Congénitas do Centro Hospitalar Universitário São João, Porto, Portugal
- ⁶Departamento de Ginecologia, Obstetrícia e Pediatria, Faculdade de Medicina, Universidade do Porto, Porto, Portugal
- ⁷Public Health Environments and Society, London School of Hygiene and Tropical Medicine, London, UK
- ⁸Bioethics Institute Ghent, Department of Philosophy and Moral Sciences Ghent University, Ghent, Belgium
- ⁹Departamento de Ciências da Saúde Pública e Forenses e Educação Médica, Faculdade de Medicina, Universidade do Porto, Porto, Portugal

Twitter Cláudia de Freitas @c6defreitas

Acknowledgements The authors would like to thank the Foundation for Science and Technology/European Regional Development Fund for providing funding for this research. One Author (ELT) of this publication is a member of the European Reference Network for Rare Hereditary Metabolic Disorders (MetabERN) - Project ID No 739 543.

Contributors CDF, SS and HM conceived the project and wrote the proposal for funding. CDF, SS, MA, ELT and MJB contributed to the development of the study design. AR and VP are advisers to the project and supported the research design. CDF is principal investigator of the project and prepared the first draft of the manuscript. All authors critically reviewed previous versions of the manuscript and approved the final version.

Funding This work is funded by FEDER through the Operational Programme for Competitiveness and Internationalisation and national funding from the Foundation for Science and Technology – FCT (Portuguese Ministry of Science, Technology and Higher Education) (Ref. POCI-01-0145-FEDER-032194), under the project 'Public and patient involvement in health data governance: a people-centred approach to data protection in genetic diseases' (Ref. FCT PTDC/SOC-SOC/32194/2017) and the Unidade de Investigação em Epidemiologia - Instituto de Saúde Pública da Universidade do Porto (EPIUnit) (Ref. UIDB/04750/2020), the individual contract grant DL57/2016/CP1336/CT0001 (CDF) and the individual contract grant IF/01674/2015 (SS).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID id

Cláudia de Freitas <http://orcid.org/0000-0002-1828-8642>

REFERENCES

- 1 Baker DJ. Genomics England: the future of genomic medical diagnosis: governmental scale clinical sequencing and potential

- walled-garden impact on global data sharing. In: Lambert CG, Baker DJ, Patrinos GP, eds. *Human genome informatics*. London: Academic Press, 2018: 269–92.
- 2 Prainsack B, Buys A. *Solidarity in biomedicine and beyond*. Cambridge: Cambridge University Press, 2017.
 - 3 European Commission. *Support for the setting-up of registries of patients affected by rare diseases available for all the ERNs*, 2019.
 - 4 Aitken M, de St Jorre J, Pagliari C, et al. 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* 2016;17:73.
 - 5 OECD. *Health data governance: privacy, monitoring and research – policy brief*. Paris: OECD Publishing, 2015.
 - 6 OECD. *Recommendation of the Council on health data governance, OECD/LEGAL/0433*. Paris: OECD Publishing, 2016.
 - 7 Hoeyer K, Tupasela A, Rasmussen MB. Ethics policies and ethics work in cross-national genetic research and data sharing: flows, nonflows, and overflows. *Sci Technol Hum Values* 2017;42:381–404.
 - 8 Kaye J, Terry SF, Juengst E, et al. Including all voices in international data-sharing governance. *Hum Genomics* 2018;12:13.
 - 9 Flores M, Glusman G, Brogaard K, et al. P4 medicine: how systems medicine will transform the healthcare sector and society. *Per Med* 2013;10:565–76.
 - 10 Benson M. Clinical implications of omics and systems medicine: focus on predictive and individualized treatment. *J Intern Med* 2016;279:229–40.
 - 11 Majumder MS, Santillana M, Mekar SR, et al. Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015–2016 Colombian Zika virus disease outbreak. *JMIR Public Health Surveill* 2016;2:e30.
 - 12 Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol* 2016;13:350–9.
 - 13 Prainsack B. The political economy of digital data: introduction to the special issue. *Policy Stud* 2020;41:439–46.
 - 14 Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. *Health Technol* 2017;7:351–67.
 - 15 Sterckx S, Rakic V, Cockbain J, et al. "You hoped we would sleep walk into accepting the collection of our data": controversies surrounding the UK care.data scheme and their wider relevance for biomedical research. *Med Health Care Philos* 2016;19:177–90.
 - 16 Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: experiences of public engagement in the Scottish health informatics programme. *Sci Public Policy* 2016;43:713–23.
 - 17 van Staa T-P, Goldacre B, Buchan I, et al. Big health data: the need to earn public trust. *BMJ* 2016;354:i3636.
 - 18 General Data Protection Regulation (GDPR). Regulation of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC. OJ 2016.
 - 19 Lochmüller H, Torrent I, Farnell J, Le Cam Y, Torrent I, Farnell J, et al. The International rare diseases research Consortium: policies and guidelines to maximize impact. *Eur J Hum Genet* 2017;25:1293–302.
 - 20 de Freitas C, Dos Reis V, Silva S, et al. Public and patient involvement in needs assessment and social innovation: a people-centred approach to care and research for congenital disorders of glycosylation. *BMC Health Serv Res* 2017;17:682–93.
 - 21 Choquet R, Maaroufi M, de Carrara A, et al. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. *J Am Med Inform Assoc* 2015;22:76–85.
 - 22 European Commission. *3rd EU health programme. call for project under the annual work programme*, 2019.
 - 23 Prainsack B, Buys A. Thinking ethical and regulatory frameworks in medicine from the perspective of solidarity on both sides of the Atlantic. *Theor Med Bioeth* 2016;37:489–501.
 - 24 Kleiderman E, Knoppers BM, Fernandez CV, et al. Returning incidental findings from genetic research to children: views of parents of children affected by rare diseases. *J Med Ethics* 2014;40:691–6.
 - 25 Council of the European Union. *Council conclusions on the transition towards an open science system*, 2016.
 - 26 World Medical Association. *Wma declaration of Taipei on ethical considerations regarding health databases and biobanks*, 2016.
 - 27 Kaye J. The tension between data sharing and the protection of privacy in genomics research. *Annu Rev Genomics Hum Genet* 2012;13:415–31.
 - 28 Blasimme A, Fadda M, Schneider M, et al. Data sharing for precision medicine: policy lessons and future directions. *Health Aff* 2018;37:702–9.
 - 29 Baía I, de Freitas C, Samorinha C, et al. Dual consent? donors' and recipients' views about involvement in decision-making on the use of embryos created by gamete donation in research. *BMC Med Ethics* 2019;20:90.
 - 30 Buyx A, Del Savio L, Prainsack B, et al. Every participant is a Pi. citizen science and participatory governance in population studies. *Int J Epidemiol* 2017;46:dyw204–84.
 - 31 Murtagh MJ, Blell MT, Butters OW, et al. Better governance, better access: practising responsible data sharing in the METADAC governance infrastructure. *Hum Genomics* 2018;12:24.
 - 32 Beier K, Schweda M, Schickantz S. Taking patient involvement seriously: a critical ethical analysis of participatory approaches in data-intensive medical research. *BMC Med Inform Decis Mak* 2019;19:90.
 - 33 Oliver K, Kothari A, Mays N. The dark side of coproduction: do the costs outweigh the benefits for health research? *Health Res Policy Syst* 2019;17:33.
 - 34 de Freitas C, Martin G. Inclusive public participation in health: policy, practice and theoretical contributions to promote the involvement of marginalised groups in healthcare. *Soc Sci Med* 2015;135:31–9.
 - 35 Renedo A, Marston C. Spaces for citizen involvement in healthcare: an ethnographic study. *Sociology* 2015;49:488–504.
 - 36 Aitken M, Tully MP, Porteous C, et al. Consensus statement on public involvement and engagement with data-intensive health research. *Int J Popul Data Sci* 2019;4:06.
 - 37 Majumder MA, Bollinger JM, Villanueva AG, et al. The role of participants in a medical information commons. *J Law Med Ethics* 2019;47:51–61.
 - 38 Escobar-Ballesta M, García-Ramírez M, De Freitas C. Taking stock of Roma health policies in Spain: lessons for health governance. *Health Policy* 2018;122:444–51.
 - 39 Vayena E, Blasimme A. Biomedical big data: new models of control over access, use and governance. *J Bioeth Inq* 2017;14:501–13.
 - 40 Fredriksson M, Tritter JQ. Disentangling patient and public involvement in healthcare decisions: why the difference matters. *Sociol Health Illn* 2017;39:95–111.
 - 41 Creswell J. *A Concise introduction to mixed methods research*. Los Angeles: Sage Publications, Inc, 2015.
 - 42 Guest G, Bunce A, Johnson L. How many interviews are enough? an experiment with data saturation and variability. *Field Methods* 2006;18:59–82.
 - 43 Strauss A, Corbin J. *Basics of qualitative research*. Thousand Oaks: Sage Publications, 1998.
 - 44 Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant* 2018;52:1893–907.
 - 45 Patton MQ. *Qualitative research and evaluation methods*. Los Angeles: Sage Publications, 2015.
 - 46 Vollmar HC, Ostermann T, Redaelli M. Using the scenario method in the context of health and health care—a scoping review. *BMC Med Res Methodol* 2015;15:89.
 - 47 Verschuuren M, Hilderink HBM, Vonk RAA. The Dutch public health foresight study 2018: an example of a comprehensive foresight exercise. *Eur J Public Health* 2019;30:30–5.
 - 48 Mohiddin A, Duggan M, Marsh S. The use of future scenario thinking for child public health in a local authority. *J Public Health* 2020;fdaa130.
 - 49 Schoemaker CG, van Loon J, Achterberg PW, et al. The public health status and foresight report 2014: four normative perspectives on a healthier Netherlands in 2040. *Health Policy* 2019;123:252–9.
 - 50 Yin RK. *Case study research design and methods*. 5 edn. Thousand Oaks: Sage Publications, 2014.
 - 51 Hannigan A, Basogomba A, LeMaster J, et al. Ethnic minority health in Ireland-Co-creating knowledge (EMH-IC): a participatory health research protocol. *BMJ Open* 2018;8:e026335.
 - 52 Verhenneman G, Claes K, Derèze JJ, et al. How GDPR enhances transparency and fosters pseudonymisation in academic medical research. *Eur J Health Law* 2020;27:35–57.
 - 53 Martin GP, Citizens MGP. Citizens, publics, others and their role in participatory processes: a commentary on Lehoux, Daudelin and Abelson. *Soc Sci Med* 2012;74:1851–3.