

Sharing data for future research—engaging participants' views about data governance beyond the original project: a DIRECT Study

Nisha Shah, MSc¹, Victoria Coathup, PhD¹, Harriet Teare, DPhil¹, Ian Forgie, PhD², Giuseppe Nicola Giordano, PhD³, Tue Haldor Hansen, MD, PhD⁴, Lenka Groeneveld, MSc⁵, Michelle Hudson, BSc⁶, Ewan Pearson, MB BChir, PhD⁷, Hartmut Ruetten, MD, PhD⁸ and Jane Kaye, LLB, DPhil¹

Purpose: Biomedical data governance strategies should ensure that data are collected, stored, and used ethically and lawfully. However, research participants' preferences for how data should be governed is least studied. The Diabetes Research on Patient Stratification (DIRECT) project collected substantial amounts of health and genetic information from patients at risk of, and with type II diabetes. We conducted a survey to understand participants' future data governance preferences. Results will inform the postproject data governance strategy.

Methods: A survey was distributed in Denmark, Sweden, The Netherlands, and the United Kingdom.

Results: In total 855 surveys were returned. Ninety-seven percent were supportive of sharing data postproject, and 90% were happy to share data with universities, and 56% with commercial companies. The top three priorities for data sharing were highly secure database, DIRECT researchers to monitor data used by other

researchers, and researchers cannot identify participants. Respondents frequently suggested that a postproject Data Access Committee should involve a DIRECT researcher, diabetes clinician, patient representative, and a DIRECT participant.

Conclusion: Preferences of how data should be governed, and what data could be shared and with whom varied between countries. Researchers are considered as key custodians of participant data. Engaging participants aids in designing governance to support their choices.

Genetics in Medicine (2019) 21:1131–1138; <https://doi.org/10.1038/s41436-018-0299-7>

Keywords: data sharing; data governance; type II diabetes; patient involvement; ethics

INTRODUCTION

The recent drive to create biobanks and data repositories to store genetic, phenotypic, and other health-related data generated from and for biomedical research is intended to encourage and facilitate data sharing for secondary research purposes.^{1–3} Researchers are increasingly prompted by funders, research organizations, and journals to make their data easily accessible.^{3,4} Furthermore, data sharing helps to optimize investment from public funds and strengthen the statistical power and rigor of research.^{2,4–11} Researchers have an ethical duty to make better use of data and avoid unnecessary repetition, particularly in genetic studies, which extract human biological samples for analysis.^{1,3,12} While

sharing research data is encouraged, researchers have an obligation to meet rigorous ethical and regulatory requirements to provide accurate and timely data, and to safeguard participants' privacy and confidentiality.^{12–16} These requirements have been strengthened by the European Union (EU) General Data Protection Regulation¹⁷ (GDPR) that came into effect across Europe in May 2018.

A key challenge is that funding bodies do not provide resources to sustain the governance of data sharing; therefore, the onus is on research projects to develop and implement data sharing policies^{2,18} that conform to legislation throughout the data collection phase and thereafter. Data access and security are tasks typically managed throughout the duration of a study.

¹HeLEX Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK; ²Division of Molecular and Clinical Medicine, University of Dundee, Dundee, UK; ³Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden; ⁴The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; ⁵Department of Epidemiology and Biostatistics, Vrije Universiteit Medical Centre, Amsterdam, The Netherlands; ⁶NIHR Exeter Clinical Research Facility, University of Exeter Medical School, Exeter, UK; ⁷Division of Population Health & Genomics, School of Medicine, University of Dundee, Dundee, UK; ⁸Translational Medicine & Early Development, Sanofi-Aventis Deutschland GmbH, Frankfurt, Germany. Correspondence: Nisha Shah (Nisha.shah@dph.ox.ac.uk)

Submitted 16 May 2018; accepted: 30 August 2018
Published online: 28 September 2018

Maintaining the same level of oversight after project completion is challenging without adequate governance and funding in place. In response to this, federal or centralized databases such as the database of Genotypes and Phenotypes (dbGaP)¹⁹ and European Genome-Phenome Archive (EGA)²⁰ have pioneered submission and access policies for management of open and controlled genome study data sets. Also, the Public Population Project in Genomics and Society (P3G) is a consortium providing policy support to the international scientific community about how data sharing can be facilitated and managed.¹³ While these mechanisms are useful and well formulated, they may not be appropriate for some research, which has collected diverse types of data and has specific priorities for data sharing and monitoring.^{4,13,18,21} Consulting researchers' and scientists' views about data sharing governance and practices are important to understand how data should be managed after the end of research projects. Importantly, research participants should also be given the opportunity for involvement in data sharing decisions, including which aspects of study data should be safeguarded and to indicate their preferred level of control over data sharing.^{10,14,22} Not engaging creates a lack of connection with the participants and reduces their control over their data being shared for future research.²³

There is growing literature about public and research participants' engagement in data sharing decision making,^{7,23–27} though this has mainly focused on what data donors' preferences would be in hypothetical contexts. Haga and O'Daniel²⁸ found that public concerns about data sharing focused on privacy and mistrust in how samples are used and who has access. McGuire *et al.*²⁹ investigated consent for data sharing and reported that research participants were happy to share genetic data with restricted access, indicating a disconnect between open data sharing policies and participants' preferences for privacy and autonomy. Finally, Ludman *et al.*³⁰ found that participants' views about submitting data into dbGaP were dependent on their ability to have control over sharing their data, and that other methods such as opt-out or notification-only for public release were less acceptable than reconsenting for submission of their genomic data and having restricted or controlled access.

These studies have mostly been conducted in North America and have focused largely on improving participation rates and the initial consent process rather than understanding participants' views about how data access should be reviewed and managed when studies end. The current study provided an opportunity to gain perspectives about future data access and governance from participants in European countries that are economically and culturally different.

Here, we report the results of a survey conducted with participants recruited to the Diabetes Research on Patient Stratification (DIRECT) Study, a European project funded by Innovative Medicines Initiative Joint Undertaking (IMI-JU).³¹ DIRECT is a European consortium of 25 partners from 10 countries, which recruited patients with and at high risk of developing type II diabetes (T2D).

As part of the DIRECT project, vast amounts of health and genetic data have been collected with the aim to identify biomarkers for the development, progress, and treatment responses of patients with and without T2D. A major focus for the project was the generation of a secure database so researchers in the consortium could access data easily for analysis. Significant effort was devoted to devising a data access policy that would allow secure and fair sharing between consortium members, with strict rules about how that would be supported and what was permissible. However, this did not extend to the postproject use of the data because it was not clear at the beginning of the DIRECT 7-year project what would be required. This raises the question about whose responsibility it would be to manage access to the data set and ensure that data is shared legally and ethically in future. Although research participants had consented to the broad use of their data postproject from the outset, it was decided that it would be ethical to elicit the views of participants about the details of the future governance structure. This aligned with the increased legal requirements under the GDPR, which at the time had not come into force. We conducted an engagement survey about data sharing governance with participants to explore the following: the importance of data access governance factors; preferences for which data types may be shared and with whom; and who should be involved in managing data access beyond the project.

MATERIALS AND METHODS

Study design and recruitment

A cross-sectional survey was conducted on a subset of the wider project sample between September 2015 and March 2016. In total, 7264 participants were enrolled into at least one of seven studies of the DIRECT project. Details of the patient cohort is described elsewhere.³¹ The study sample consisted of patients diagnosed with T2D and individuals at high risk of the disease but who were receiving no treatment for diabetes. From the participants enrolled into the project, a subset of 1082 were approached through ten specialized diabetes clinics and university study centers across four of the collaborating countries including the United Kingdom, Denmark, Sweden, and The Netherlands. Survey respondents were adults aged 18 to 80, of white European descent, and had already given a broad consent for data to be collected, stored, and used by both DIRECT researchers and researchers in the wider scientific community, who had been granted access.

Participants were recruited in one of two ways. First, during their follow-up appointments, as part of the main DIRECT study. Upon arrival for their appointment, they were given a blank envelope containing the survey, an information sheet, and invitation letter. If participants decided to take part, they were instructed to place the completed survey inside the blank envelope, seal it, and place it in a designated collection box. Second, participants who had already completed all their DIRECT study visits were mailed the invitation letter, information sheet, and survey, with a prepaid envelope for

return. Consent was implied with the return of a completed survey.

Survey development

The survey underwent an iterative development process, through review with wider consortium members, specialist diabetes nurses, and T2D patients. They provided feedback about the survey's content, questions, and structure. The final survey was composed of three sections with 24 items (see supplementary information survey): section 1 comprised of questions relating to respondents' motivations to take part in medical research and their experiences since being involved in the DIRECT project; section 2 related to opinions about data sharing; and section 3 collected respondents' socio-demographic information. Response categories varied by item and included multiple-choice answers, as well as 5-point Likert scales assessing happiness, importance, and support categories for given statements. Items about risks and benefits to sharing genetic information and attitudes toward privacy and advancing research were used from an existing survey.³² Results reported here relate to questions specifically about support for sharing data; what and with whom data can be shared, and how it should be managed; and sociodemographic characteristics. The study was approved by institutional review boards in Denmark (The Secretariat of the Scientific Ethics Committees for the Capital Region Protocol no. H-1-2011-166 Note no. 50965, and H-1-2012-100 Note no. 50694), Sweden (Regional Ethics Examination Board in Lund Dnr 2015/815 and Dnr 2015/843), The Netherlands (Medical Ethics Review Committee Vrije Universiteit Medical Centre Protocol 2012.222), and the United Kingdom (Newcastle and North Tyneside 1 Research Ethics Committee 12/NE/0132; East of Scotland Research Ethics Service 11/ES/0046; and 12/ES/0034).

Data analysis

Descriptive summaries of responses were calculated as frequencies and percentages, and results stratified by country are provided in the supplementary materials. All variables in the analyses were categorical. The Likert scale for item "How do you feel about your nonidentifiable data being shared for medical research?" was collapsed from 5 points into binary responses "Supportive" versus "Not supportive." Response options in the categorical variables were collapsed due to small numbers in the extreme options, and to balance the distributions as much as possible. Extreme options were grouped but middle categories were retained to minimize the information lost through collapsing. The "I don't know" and "Prefer not to say" options were treated as missing because of minimal (less than 5%) or zero counts. Chi-square analyses compared sociodemographic characteristics stratified by country. Univariate binary logistic regressions were conducted to assess associations between sociodemographic characteristics and support for data sharing, and multivariate binary logistic regressions were conducted for estimates of support for sharing data in future research by sociodemographic characteristics. The explanatory variables entered into the

regression models were age, gender, country, education level, self-rated knowledge of genetics, diabetes status, previously worked in health or medicine, and self-reported health. Results are presented as odds ratios (ORs) with 95% confidence interval (CI) and significance level $p \leq 0.05$. Due to small percentages of missing data, only complete cases were analyzed for the results presented. Internal consistency (Cronbach's α) is reported for the following item scales: happiness to share different data types, happiness to share with different research groups, and the importance of data governance factors (see Tables 3 and 4). All analyses were performed using SPSS 22 (SPSS, Inc., Chicago, IL).

RESULTS

Demographics and response rate

Completed surveys were received from participants attending clinic follow-up visits and by post. A total of 855 surveys were returned. The combined response rate for the United Kingdom, Denmark, and Sweden was 79%. The highest response came from The Netherlands (94%), United Kingdom (86%), followed by Denmark (70%), and the smallest from Sweden (59%). Nearly three-quarters of respondents were aged 61 or over (73%), 43% were female, 41% with vocational or professional qualifications, compared with 19% with degree level, and 37% with secondary education. Seventy percent had been diagnosed with T2D at the time of completing the survey (Table 1). When asked "How would you rate your knowledge of genetics?" 837 respondents rated their knowledge on average as fair (median = 3), where 1 = very good, 2 = good, 3 = fair, 4 = poor, 5 = very poor.

Support for data sharing

Participants were asked about their level of support to share their de-identified data for medical research after the project had ended. The survey found that 97% were either very or fairly supportive of their data being shared. Country of origin was independently associated with support for sharing data ($X^2 = 11.58$, $df = 3$, $p = 0.009$; see supplementary Table S1 for descriptive results and Table S2 for the univariate results), but there were no other significant associations between other sociodemographic factors and support for data sharing. After adjusting for all other factors, Dutch respondents were significantly less likely to support sharing of their de-identified records for future research compared with UK respondents (OR = 0.211; 95% CI, 0.073 to 0.608; $p = 0.004$). Respondents with vocational or professional qualifications were more likely to support data sharing compared with those with high school (or lower) education (OR = 3.82; 95% CI, 1.243 to 11.734; $p = 0.019$) (Table 2). Other education level groups were not significantly associated with support for data sharing.

Level of happiness for sharing different types of data and with different research groups

Respondents were asked about their level of happiness to share different types of data (Table 3). More than 85% rated

Table 1 Participant characteristics by country^a

Characteristics	All respondents (n=855) ^b	United Kingdom (n=422)	Denmark (n=265)	Sweden (n=56)	The Netherlands (n=112)	P ^c
Gender						
Male	481 (57.3)	249 (60.6)	127 (48.1)	35 (64.8)	70 (63.6)	0.003
Female	358 (42.7)	162 (39.4)	137 (51.9)	19 (35.2)	40 (36.4)	
Age (years)						
≤60	228 (27.2)	115 (28.0)	71 (26.9)	9 (16.7)	33 (30.0)	0.268
61–70	395 (47.1)	185 (45.1)	129 (48.9)	34 (63.0)	47 (42.7)	
≥71	215 (25.7)	110 (26.8)	64 (24.2)	11 (20.3)	30 (27.3)	
What is your highest level of education?						
≤Secondary/high school	300 (36.5)	174 (43.4)	46 (17.7)	38 (71.7)	42 (39.3)	<0.001
Vocational/professional	336 (40.9)	139 (34.7)	135 (51.9)	8 (15.1)	54 (50.4)	
≥Bachelor's degree	159 (19.3)	79 (19.7)	68 (26.2)	5 (9.4)	6 (5.6)	
Other	27 (3.3)	9 (2.2)	11 (4.2)	2 (3.8)	5 (4.7)	
Diagnosed with T2D						
Yes	589 (70.3)	408 (99.0)	37 (14.0)	52 (96.3)	92 (85.2)	<0.001
No and unsure	249 (29.7)	4 (1.0)	227 (86.0)	2 (3.7)	16 (14.8)	
Have you had a job in medicine/health?						
Yes	169 (20.12)	89 (21.5)	54 (20.6)	10 (18.5)	16 (14.5)	0.433
No	671 (79.88)	325 (78.5)	208 (79.4)	44 (81.5)	94 (85.5)	
How would you rate your own health?						
Very good or good	529 (63.3)	235 (57.5)	187 (71.4)	38 (70.4)	69 (62.7)	<0.001
Fair	252 (30.2)	133 (32.5)	70 (26.7)	14 (25.9)	35 (31.8)	
Very poor or poor	54 (6.5)	41 (10)	5 (1.9)	2 (3.7)	6 (5.5)	
How would you rate your genetics knowledge?						
Very good or good	130 (15.5)	55 (13.3)	43 (16.4)	12 (23.0)	20 (18.3)	0.36
Fair	381 (45.5)	186 (44.9)	125 (47.7)	20 (38.5)	50 (45.9)	
Very poor or poor	326 (38.9)	173 (41.8)	94 (35.9)	20 (38.5)	39 (35.8)	

T2D type 2 diabetes.

^aNot all participants answered all questions and "Prefer not to say" option in all questions not included due to small cell counts.^bN(%).^cPearson's Chi-square.

being happy or very happy to share medical history, genetic information blood test results, and lifestyle information, compared with 64% of respondents being happy to share their personal information.

Respondents were asked to rate their level of happiness to share study data with different research groups (Table 3). Overall, respondents were most happy to share data with research teams in universities across Europe (90% happy or very happy), and least happy to share with commercial companies such as drug companies (56%). While there is support to share data across different types of research groups, our findings show a moderate willingness to do so. Very few respondents indicated being "very happy" to having their data shared across groups, though these were more than those who were "very unhappy." The percentage of "happiness" to share with different research groups, stratified by country, is outlined in supplementary Figure S1.

Data governance and data access committee preferences

Respondents were asked how important the listed data governance factors were when their data are being shared with other research teams. The factors rated as highest importance were "The database is highly secure" (89%; extremely important and fairly important), followed by "Members of the DIRECT project can monitor how my data is being used" (74%). For all data governance factors, the median value = 2 (where 1 = Extremely important, 2 = Fairly important, 3 = Neither important or unimportant, 4 = Fairly unimportant, and 5 = Not at all important). This indicated that respondents believed that governance factors were paramount; for example, that all data access requests must be reviewed by appointed experts, and that someone from the DIRECT consortium should monitor use of data (Table 4).

When asked who should be involved in any DIRECT poststudy Data Access Committee (DAC), respondents

Table 2 Multivariate binary logistic regression of factors associated with support for sharing de-identified data to be shared for medical research^{a,b} (N = 798)^c

	OR	95% CI	p
Gender			
Female	REF		
Male	0.81	0.32–2.07	0.66
Age			
≤60	REF		
61–70	1.29	0.44–3.73	0.64
≥71	1.25	0.38–4.07	0.71
Country			
United Kingdom	REF		
Denmark	0.38	0.07–2.19	0.28
Sweden	0.43	0.1–1.74	0.24
The Netherlands	0.21	0.07–0.62	0.004
What is your highest level of education?			
≤Secondary or high school	REF		
Vocational or professional	3.77	1.23–11.58	0.02
≥Bachelor’s or higher	4.00	0.8–20.06	0.09
Other	1.89	0.2–17.56	0.57
Diagnosed with T2D^d			
Yes	REF		
No	5.10	0.76–34.26	0.09
Have you had a job in medicine or health?			
Yes	REF		
No	2.57	0.9–7.38	0.08
How would you rate your own health?			
Very good or good	REF		
Fair	1.22	0.44–3.38	0.70
Very poor or poor	0.82	0.16–4.26	0.81
How would you rate your genetics knowledge?			
Very good or good	REF		
Fair	1.46	0.49–4.39	0.50
Very poor or poor	2.12	0.61–7.37	0.24

CI confidence interval, OR odds ratio, T2D type 2 diabetes.

^aReference = unsupportive.

^bAll variables entered into the model were categorical in nature.

^cNot all respondents answered all questions, so complete cases included in model.

^dDue to small cell counts “I don’t know” answers grouped with “No.”

selected from multiple options. Figure S2 (supplementary information) summarizes those who respondents considered important for the DAC. The most frequently selected type of person or group to be involved were a DIRECT researcher (687 counts), diabetes doctor or nurse (582), diabetes patient representative (496), DIRECT participants (268), lawyer (212), and lay person (116). There were nine selections for “Other,” however, the survey did not ask for respondents to provide further information.

DISCUSSION

Eliciting views of study participants at the individual or group level about whether, how, and with whom their data should be shared after a study has ended could help inform future data

Table 3 Happiness to share different types of data and with different groups for medical research^a

	Total ^{b,c}	Very unhappy	Unhappy	Neither unhappy nor happy	Happy	Very happy
Happiness to share different data types (α = 0.84)^d						
Medical history	842	0 (0)	32 (3.8)	89 (10.6)	689 (81.8)	32 (3.8)
Genetic information	838	1 (0.1)	24 (2.9)	84 (10)	695 (82.9)	34 (4.1)
Blood test results	838	0 (0)	15 (1.8)	61 (7.3)	725 (86.5)	37 (4.4)
Information about lifestyle (e.g., diet, exercise, smoking)	837	2 (0.2)	28 (3.3)	90 (10.8)	685 (81.8)	32 (3.8)
Personal information (e.g., marital status, age, income status)	832	4 (0.5)	131 (15.7)	163 (19.6)	506 (60.8)	28 (3.4)
Happiness to share data with these groups (α = 0.92)^d						
Research teams in European universities	839	0 (0)	26 (3.0)	60 (7.2)	718 (85.6)	35 (4.2)
Research teams in global universities	837	1 (0.1)	36 (4.3)	74 (8.8)	691 (82.6)	35 (4.2)
Government funded organizations involved in research	839	2 (0.2)	26 (3.1)	79 (9.4)	699 (83.3)	33 (3.9)
Commercial research companies (e.g., drug companies)	829	6 (0.7)	147 (17.7)	209 (25.2)	449 (54.2)	18 (2.2)
Charities involved in research	830	1 (0.1)	80 (9.6)	163 (19.6)	554 (66.8)	32 (3.9)
Patient organizations involved in research	831	1 (0.1)	49 (5.9)	128 (15.4)	623 (75)	30 (3.6)

^a Respondents were asked “Would you be happy if the following types of data were shared with other researchers?” and “Would you be happy for your data (nonidentifiable) to be shared with the following groups of people?”

^b Not all respondents answered these statements.

^c N (%) responses.

^d Cronbach’s alpha coefficient assessing internal consistency of the scales.

Table 4 Importance of data governance factors when data is shared^{a, b}

	Total ^c	Extremely important	Fairly important	Neither important or unimportant	Fairly unimportant	Not at all important
Importance of data governance factors ($\alpha = 0.88$)^d						
Researchers cannot identify me	828	28 (3.4)	549 (66.3)	139 (16.8)	104 (12.6)	8 (0.9)
The database is highly secure	829	50 (6)	689 (83.1)	40 (4.8)	45 (5.4)	5 (0.6)
I can withdraw my data at any time	826	20 (2.4)	480 (58.1)	199 (24.1)	114 (13.8)	13 (1.6)
An expert committee must approve every application before granting access	828	19 (2.3)	569 (68.7)	147 (17.8)	86 (10.4)	7 (0.8)
Members of the DIRECT project can monitor how my data is being used by other research teams	822	29 (3.5)	587 (71.4)	144 (17.5)	59 (7.2)	3 (0.4)

DIRECT Diabetes Research on Patient Stratification project.

^aRespondents were asked "How important are the following factors when your data is being shared with other research teams outside of the *DIRECT* database?"

^bNot all respondents answered these statements.

^c*N* (%) responses.

^dCronbach's alpha coefficient assessing internal consistency of the scale.

access decisions.^{5,10} This study preceded the GDPR. However, it aligns with the legal requirements for the use, storage, and management of data that have undergone a significant change. The GDPR strengthens individuals' rights and imposes greater obligations of accountability and transparency on researchers. New protections for individuals are the right to object to how their data is processed, the right to withdraw consent, and the right to be contacted in case of data breaches (Art. 7[3] and 34) (ref. ¹⁷). To be able to exercise these rights participants must receive information about the use of the data. Beyond the GDPR, offering participants' engagement in data management decisions further strengthens transparency and accountability.

While many studies have investigated the choices participants would hypothetically make to inform future data sharing for research policy and practice,^{28,30} this study engaged with participants to help develop the postproject data sharing strategy for the data obtained from their participation. This approach is not commonly found in the literature. Participants had undergone various data collection phases and were familiar with how the research was conducted and what data were collected; they were well suited to provide insights on if and how their data should be shared. Further qualitative research may provide insights about the reasons for respondents' selections. Ongoing engagement and involvement is also key if research participants are to become integral stakeholders in data sharing governance.

The survey results indicate that overall respondents were supportive of sharing their de-identified data outside the consortium. This result mirrors North American studies investigating perspectives about de-identified data sharing, and reiterates the conditions under which respondents are prepared to allow data to be shared, which included that data are held and disseminated securely,²⁸ donors have control over what data are shared,²⁵ donors should be asked if their data can be shared,¹² and that data are shared for the benefit of wider society.²³ Interestingly, Dutch respondents were significantly less likely to be supportive of sharing their de-

identified data compared to UK respondents. Similarly, a special Eurobarometer survey investigated European citizens' attitudes toward the impact of digitization in daily life, and asked about their willingness to share anonymized data with different research groups.³³ It reported that Dutch respondents were less willing than Swedish and Danish respondents to share anonymized data with public authorities and public sector companies for medical research.³³ Further investigation is needed to understand why people in The Netherlands appear to be more protective of their data.

The survey set out to determine which types of data respondents were content to share outside of the *DIRECT* consortium and to consider whether their views varied by data type. The results indicated that respondents were moderately happy to share most types of information, with least support for sharing personal information. These findings echo previous literature investigating willingness to consent to share data.^{25,29,34} Other studies also show support for data sharing, for medical research, as long as data are de-identified.^{8-10,34}

In our sample, respondents were happier to share data with universities and least happy to share with commercial companies. The reasons for this were not explored and warrant further investigation. In comparison with these results, it has been suggested that public concerns relating to sharing data with government or the pharmaceutical industry were centered around beliefs about past reported misuses of public data, specifically of vulnerable and minority communities.²⁸ Additionally, McGuire *et al.*³⁵ reported that participants in focus groups, investigating data sharing from genome-wide association studies, entrusted home institutions and local investigators to protect privacy and their data relative to federal control. Similarly, a report commissioned by the Wellcome Trust in the UK²⁶ found that willingness to share data is influenced by trust in the institution and the extent to which patients are informed about who their data are being shared with, and what type of data, particularly in relation to commercial entities.

The survey was a mechanism for involving DIRECT participants in response to growing research about providing data donors with the choice to share and exercise control over their data; data privacy was an important aspect of their consent to share data, particularly for use in future research and the potential of for-profit organizations to access their data.¹⁰ It has been claimed there is a need, however, to redress the balance between privacy and openness of data that may be sensitive in nature, such as in genomic studies;¹⁵ engaging with stakeholders to determine their recommendations on how to proceed may facilitate this. Kim *et al.*²³ found that patients who believed that sharing health-care data through specific health-care networks, which were considered secure and already protected privacy, were more likely to consent to share. Privacy and security are high priorities for patients and research participants, particularly when it is proposed that data may be shared with commercial companies²⁶ and governmental organizations.^{28,30} The likelihood that patients and participants consent to use of their data is dependent on public trust in gatekeepers and the potential users of personal data. It has been argued that retention and use of existing data is more ethical and less wasteful,³⁶ not least because it eliminates the burden of participants providing data repeatedly, provided participants are supportive of such practices and trusting of the research initiative. Involving patients and public in research data access and governance decisions facilitates transparency and trustworthiness of practices and is in line with the GDPR.

This study found that respondents are happy for their data to be shared beyond the original research, so long as measures are in place to protect them and provide control. However, that is increasingly difficult to do if there are no resources specifically allocated for this. Leadership from funding bodies should be key in enabling harmonization of data access and sharing.^{4,5,37} Shabani *et al.*^{5,37} stated that funding bodies should lead the way in which data access is arranged and that the creation of centralized DACs should be considered. In the absence of this support, databases such as dbGaP and EGA have paved the way to build mechanisms for data management and sharing. However, such guidance may be too generalized and broad for specific research projects, prompting individual research bodies to develop localized strategies relevant to their original research.^{4,13,18,21}

DIRECT is an example of a large prospective project, spanning multiple European countries, that has taken significant steps to promote data sharing in a secure environment within the consortium and wishes this to be upheld once the project has ended. How this should be implemented by projects of this type needs to be addressed within the wider scientific community, and importantly, should include the involvement of study participants. This is particularly so if researchers are committed to data being shared and reused to maximize its use while simultaneously safeguarding the privacy of data subjects.

In conclusion, considering the survey findings and supporting literature, one solution might be to form a DAC

to be responsible for data sharing and management beyond the project.^{4,16,37–40} However, as the survey results highlighted, nuances in the preferences of respondents on how data should be managed and shared would need to be accommodated within any policy and practice directives. The challenge now is to identify the mechanisms by which respondent preferences can be followed through and offer controllability. The next stage of DIRECT's data governance strategy development will be to identify appropriate persons and mechanisms to be responsible for data sharing and management, identify repositories to store data, and consider how to facilitate DIRECT participants' involvement in all future data sharing decisions.

Strengths and limitations

This is a novel large European cross-sectional study to understand data sharing preferences of patients enrolled in the DIRECT project. It was important to investigate participants' perspectives from different European countries to enable development of data governance strategies relevant for this region. While this was a homogeneous self-selecting sample, the survey was the first step toward engagement with study participants to incorporate their views about governance of their data in the future. Although there was a large response rate, several limitations existed. The nature of self-reported studies introduces response bias in that it is difficult to establish whether respondents answered truthfully. However, to minimize this, participants were informed in the study materials that their information would be kept confidential and that they would remain anonymous. While the reliability of the item scales was within the acceptable range, the initial piloting of the survey only elicited content validity, and further validation is required. The average age of respondents was older than the general population, as T2D is more prevalent in people middle-aged and above. Additional demographic data such as comorbidities, occupation, and income were not collected on this occasion. The eligibility criteria for the DIRECT project did not include patients from ethnicities other than white European descent, due to the particular design and aims of the wider project. The sample therefore may not be representative of the overall populations in the countries investigated and precludes any conclusion regarding data governance attitudes more widely. Future research could replicate this study in different cross-cultural groups, proportionately represented in the populations of the countries involved. The survey was not designed to include open responses; it was translated into four languages and as such would have introduced complexities such as back-translation into English, which was beyond the remit of this study.

ELECTRONIC SUPPLEMENTARY MATERIAL

The online version of this article (<https://doi.org/10.1038/s41436-018-0299-7>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

The authors would like to thank all of the teams in the consortium that assisted with the survey development, distribution, and collection. The work leading to this publication has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115317 (DIRECT), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in-kind contribution.

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

- Duke C, Porter J. The ethics of data sharing and reuse in biology. *Bioscience*. 2013;63:483–489.
- Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. Data sharing in genomics—re-shaping scientific practice. *Nat Rev Genet*. 2009;10:331–335.
- Heeney C, Kerr SM. Balancing the local and the universal in maintaining ethical access to a genomics biobank. *BMC Med Ethics*. 2017;18:80.
- Shabani M, Borry P. “You want the right amount of oversight”: interviews with data access committee members and experts on genomic data access. *Genet Med*. 2016;18:892–897.
- Shabani M, Knoppers BM, Borry P. From the principles of genomic data sharing to the practices of data access committees. *EMBO Mol Med*. 2015;7:507–509.
- Harron K, Gilbert R. Research: Increasing value, reducing waste. *Lancet*. 2014;383:1124.
- Riordan F, et al. Patient and public attitudes towards informed consent models and levels of awareness of electronic health records in the UK. *Int J Med Inform*. 2015;84:237–247.
- Kaufman D, Murphy J, Scott J, Hudson K. Subjects matter: a survey of public opinions about a large genetic cohort study. *Genet Med*. 2008;10:831–839.
- Porteri C, Pasqualetti P, Togni E, Parker M. Public's attitudes on participation in a biobank for research: an Italian survey. *BMC Med Ethics*. 2014;15:81.
- Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. *Expert Rev Mol Diagn*. 2014;8:1053–1065.
- Tudur Smith C, et al. How should individual participant data (IPD) from publicly funded clinical trials be shared? *BMC Med*. 2015;13:298.
- Page SA, Manhas KP, Muruve DA. A survey of patient perspectives on the research use of health information and biospecimens. *BMC Med Ethics*. 2016;17:48.
- P3G Consortium, Church G, Heeney C, et al. Public access to genome-wide data: five views on balancing research with privacy and protection. *PLoS Genet*. 2009;5:e1000665.
- Mascalzoni D, et al. International Charter of principles for sharing biospecimens and data. *Eur J Hum Genet*. 2015;23:721–728.
- Vayena E, Gasser U. Between openness and privacy in genomics. *PLoS Med*. 2016;13:e1001937.
- Ellenberg S, Fleming T, DeMets D. Data monitoring committees in clinical trials: a practical perspective. John Wiley & Sons, Ltd.: Chichester, England 2002.
- Regulation (EU) 2016/679 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 (General Data Protection Regulation). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679>.
- Kaye J, Hawkins N. Data sharing policy design for consortia: challenges for sustainability. *Genome Med*. 2014;6:4.
- Mailman MD, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007;39:1181–1186.
- Flicek P, Birney E. The European Genome Archive: background and Implementation [White paper]. 2007. https://www.ebi.ac.uk/ega/sites/ebi.ac.uk.ega/files/documents/ega_whitepaper.pdf. Accessed 23 March 2018.
- Langhof H, Kahrs H, Sievers S, Streh D. Access policies in biobank research: what criteria do they include and how publicly available are they? A cross-sectional study. *Eur J Hum Genet*. 2017;25:293–300.
- Sanderson SC, et al. Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US. *Am J Hum Genet*. 2017;100:414–427.
- Kim K, Sankar P, Wilson M, Haynes S. Factors affecting willingness to share electronic health data among California consumers. *BMC Med Ethics*. 2017;18:25.
- Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME. Public and biobank participant attitudes toward genetic research participation and data sharing. *Public Health Genomics*. 2010;13:368–377.
- Bell EA, Ohno-Machado L, Grando MA. Sharing my health data: a survey of data sharing preferences of healthy individuals. *AMIA Annu Symp Proc*. 2014;2014:1699–1708.
- Ipsos Mori. The one-way mirror: public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. 2015. <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf>. Accessed 23 March 2018.
- D'Abramo F, Schildmann J, Vollmann J. Research participants' perceptions and views on consent for biobank research: a review of empirical data and ethical analysis. *BMC Med Ethics*. 2015;16:60.
- Haga SB, O'Daniel J. Public perspectives regarding data-sharing practices in genomics research. *Public Health Genomics*. 2011;14:319–324.
- McGuire AL, et al. To share or not to share: a randomized trial of consent for data sharing in genome research. *Genet Med*. 2011;13:948–955.
- Ludman EJ, et al. Glad you asked: participants' opinions of re-consent for dbGap data submission. *J Empir Res Hum Res Ethics*. 2010;5:9–16.
- Koivula R, et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: rationale and design of the epidemiological studies within the IMI DIRECT Consortium. *Diabetologia*. 2014;57:1132–1142.
- Oliver JM, et al. Balancing the risks and benefits of genomic data sharing: genome research participants' perspectives. *Public Health Genomics*. 2012;15:106–114.
- Eurobarometer. Attitudes towards the impact of digitisation and automation on daily life. 2017. <https://doi.org/10.2759/835661>. Accessed 9 May 2018.
- Spencer K, et al. 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*. 2016;18:e66.
- McGuire AL, et al. Ethical and practical challenges of sharing data from genome-wide association studies: the eMERGE Consortium experience. *Genome Res*. 2011;21:1001–1007.
- M Walport, P Brest. Sharing research data to improve public health. *The Lancet*. Volume 377, Issue 9765, 12–18 February 2011, Pages 537–9
- Shabani M, Dyke SOM, Joly Y, Borry P. Controlled access under review: improving the governance of genomic data access. *PLoS Biol*. 2015;13:e1002339.
- Herson J Data and safety monitoring committees in clinical trials. Chapman & Hall: Boca Raton, Florida 2009
- Tereskerz P. Data safety monitoring boards: legal and ethical considerations for research accountability. *Account Res*. 2010;17:30–50.
- Shabani M, Dove ES, Murtagh M, Knoppers BM, Borry P. Oversight of genomic data sharing: what roles for ethics and data access committees? *Biopreserv Biobank*. 2017;15:469–474.