Researcher access to mental health data: results from an online consultation

Abigail E Russell[1], Tamsin Ford[2], Andrew McIntosh[3], Peter B Jones[2], Sarah Shenow[4], Ginny Russell[1] and Sally McManus[5]

[1] University of Exeter College of Medicine and Health, UK

[2] Department of Psychiatry, University of Cambridge, UK

[3] Division of Psychiatry, University of Edinburgh, UK

[4] MQ Mental Health Research, UK

[5] City, University of London, London, UK

Corresponding author: Abigail Russell; a.e.russell@exeter.ac.uk; 01392722985

Contents

Summary

Many scientists and researchers use information from hundreds or thousands of people to understand more about health and wellbeing, and to improve treatments and support for those with health problems. This information, or data, is collected in many ways and is kept secure by organisations or individuals known as 'data controllers'. Their job is to manage the processes that allow bona fide researchers access to these data in a way that respects the privacy and confidentiality of the people the data refers to.

Since the new data protection regulations (GDPR) introduced in the UK in 2018, we have heard that researchers and scientists are finding it more difficult to access the health data they need to do important scientific research. In order to collect more information about whether people are having problems with accessing data, and to understand more about what problems people were having, we conducted a short online consultation or survey asking about data access.

We advertised the consultation by saying that MQ Mental Health Research (a charity that aims to improve mental health via research) and NHS Digital were interested in hearing about how the data access process could be improved, so it is possible that participants tended to be those who had faced difficulties. Sixty-three people took part in the survey, which was available for just over one month in late 2019. Eleven of them had not yet tried to get access to data, but said they were expecting problems if they were to try. Sixteen people were currently trying to access data, and 36 had successfully accessed data.

The 52 people who had tried to, or had accessed, data reported a wide range of problems with their applications. Often each person experienced multiple problems. The application process has lasted between 2 and 70 months (median 12 months), bearing in mind that some of these were not yet completed. There were two common issues. Firstly, the process of application was lengthy and complicated, and not clearly explained. Secondly, researchers reported getting different advice from different people within the same organisation on the same procedures. In addition, some people reported difficulty finding a person who was senior enough in their own institution to sign off the forms.

Our recommendations for data controllers, research organisations, and individual researchers following this consultation are shown in Box 1. Our consultation suggests that data access processes need to be clearer and more streamlined and applied in the same way across different data controllers. This would help to ensure that people seeking to access data get consistent advice and consistent responses. We recommend that timelines for the application process, as well as the time between application approval and data provision, are shortened. We need a better balance between data security and privacy on the one hand, and the availability of data to provide evidence for policy and practice in line with the open science movement on the other. This is particularly true where research stands to directly benefit individuals or societies. Those we consulted had a lot of suggestions for specific amendments

to parts of the existing process, which were underpinned by frustration and confusion with the process as a whole. Our findings highlight the need for rapid improvement and we recommend actions for all stakeholders within the next 12 months. Given that many other countries routinely publish studies using recent epidemiological data, failure to improve data access processes will clearly result in the UK lagging behind in the international scientific field.

Box 1. Recommendations

Based on our consultation, we make the following recommendations of actions to take to improve the situations over the next 12 months:

All stakeholders, governmental level to lead:
- Redesign a proportionate and uniform process to balance the risks of breaches of privacy with the benefit to science and health policy based on the level of sensitivity of the data (e.g. low, medium, high).

Individual data-user level:
- Identify and undertake appropriate training relating to data security and GDPR.
- Demonstrate appropriate supervision from more experienced individuals where necessary.

Institutions hosting those requiring data access:
- Maintain infrastructure to support the necessary permissions, such as an Information Governance Toolkit/Data Security and Protection Toolkit.
- Ensure that appropriately trained senior staff have sufficient time to advise as well as to review clear timelines for the sign-off applications.
- Link those applying for data access with those who have successful applications to optimise initial applications.

Data custodians:
- Ensure transparent, consistent and clear information about the access process with detailed information about what is required at each step.
- Provide additional low-risk open access datasets that can be shared with institutions for student projects to ensure that we encourage and develop the next generation of data scientists.
- Link those applying for data access with those who have successful applications to optimise initial applications.

Governmental level
- Establish an all-party parliamentary group to review how to optimise safe and legal access to data.
- Clarify the remit of data controllers, data curators and data processors.
- Identify and share examples of best practice by data controllers and organisations that facilitate the sharing of data (e.g. the ESRC UK Data Service).

Background

Epidemiologists, health researchers, students, practitioners, commissioners, policy makers and those working in the third sector often rely on anonymised research or administrative datasets in order to study how best to improve services, treatments, and the identification and prevention of health problems. Such data are collected nationally and internationally through funded research programmes, audits and national surveys. For example, research study designs include longitudinal cohorts, which follow the same individuals repeatedly over time, representative national surveys capturing the health and lifestyle of thousands of individuals and families as well as administrative data that include diagnoses and treatment contacts in primary, secondary and tertiary care. Many of these large datasets capture a huge range of information that can be used to advance knowledge in many areas, beyond the initial research purpose or the running of health (or other) services. Those who participate in survey and cohort research often commit to dedicating their time and information under the impression that their data will lead to scientific advancements to improve the health and wellbeing of others. As such, we have a responsibility to ensure these valuable resources are not wasted. Much of this initial research is already funded by the public purse or voluntary sector, as is much of the secondary analysis, so we need to minimise additional resource use in terms of researcher time or financial charges where possible.

Varied procedures exist in terms of data curation, storage and access. These procedures are partly dependent on the conditions of the original funders, data controllers, and on current data protection regulations. With the inception of the General Data Protection Regulation (GDPR) in 2018 (that is to be replaced by the highly similar UK-GDPR following the Brexit transition process), some evidence suggests that it has become more difficult for researchers in the UK to access anonymised or pseudonymised[1] datasets, particularly those regarding mental health and other health data (2). In order to ascertain the nature, range and impacts of data access difficulties, we conducted a short online consultation aimed at capturing the experiences of researchers who routinely use data. We had the aim of developing recommendations for those who control data access in order to streamline and facilitate future access procedures.

Box 2. A particular problem for mental health data? (1)

Data access problems seem particularly acute for mental health surveys, perhaps because of the sensitive nature of the topic. For example, the Office for Statistical Regulation have been conducting a systemic review of mental health statistics in England. In an update on that review (1) they report:

"Many of the users we have spoken to raised concerns about mental health data access, in particular in relation to the Adult Psychiatric Morbidity Survey (APMS) and Mental Health of Children and Young People Survey (MHCYPS). There has been a dramatic decline in use of the APMS data since DARS [*the Data Access Request Service is the department in NHS Digital that deals with requests to access health data*] became responsible for its distribution.

Figures cited in a Lancet Psychiatry editorial show that the most recent survey, conducted in 2014, which is available via DARS had been accessed by 22 unique users as of August 2019, whereas its 2007 predecessor, which is still available via UKDS, had 288 new downloads between 2016 and 2019.

The official statistics based on the MHCYPS were released in November 2018 but the data has not yet been made available to researchers, despite several groups expressing interest in secondary analyses. We have raised our concerns about this lack of timeliness with NHS Digital, highlighting that it does not align with the principles in the Code of Practice. We understand that work is underway with Information Governance colleagues to resolve this situation and we look forward to seeing this communicated to users as soon as a solution is agreed and can be actioned. We will monitor progress on MHCYPS and APMS data access as part of our mental health systemic review follow-up actions."

MQ, a key UK mental health research charity with close engagement with patient-led groups such as usemydata.org has also recognised the need to address particular issues around access to health surveys, and routine data. MQ convened a workshop, resulting in a Personal View (under review), which has also informed the recommendations in this article. However, data access concerns are not restricted to the mental health field alone (3) and our report should be read with the broader data access environment in mind.

Methods

The short online survey was programmed using Jisc online surveys, with skips and routing to minimise participant burden, and hosted by the University of Exeter. Response options included both open and closed answers, while some structured questions allowed respondents to endorse more than one potential response in recognition that researchers may have made more than one application, or had varied experiences during a single application. More than one response could be recorded for many questions. The questionnaire could be accessed online via a url. Invitations to participate were shared via social media, mostly on Twitter. Both organisational accounts (e.g. MQ) and individual accounts (e.g. of the authors) were used to encourage participation. An invitation and the questionnaire link were also emailed directly to known data-user groups, such as those based in particular university departments or third sector organisations. The wording of the invitation was informal and varied with target audience. It indicated that the MQ Data Science group and NHS Digital were interested in hearing about how the data access process could be improved, so it is possible that participants tended to be those who had faced difficulties. The consultation was introduced as being about access to administrative or survey data for heath research. The survey was active from 28th October to 2nd December 2019. 63 people took part: 11 had not applied for data; 16 had tried or were trying to access data but had not yet received any; and 36 had accessed data. To reassure participants of confidentiality and encourage candid reporting, no information was requested on participants' name, role or institutional affiliation.

Results

Consultation participants who had not applied for data

11 (17%) of the consultation participants had not started an application. All but one of these expected difficulties, and two were unable to proceed to an application stage because of them. The most common anticipated difficulties were with the overall complexity of the application process (6 out of 10 participants), lack of clarity about the process (3/10), and understanding what the requirements actually were (5/10). 40% (4/10) did not expect to have the time to be able to undertake an application. Others expected problems with communication: with the data controller (4/10), between organisations involved (4/10), and with inconsistency between views of different organisations involved (3/10). Expected problems also related to specific stages and requirements: such as registering with the Data Access Request Service (DARS) or its equivalent (4/10), others cited the Information Governance Toolkit (either their institution having required it, 3/10; or not having the registration details for it, 2/10- see below for more detail on the toolkit). Privacy notices, protocol production, demonstrating benefit, demonstrating outputs and impact, and accessing a secure room were all indicated as foreseen problems. Consequences included issues with securing funding because of funders concerns that the individual could have data access (20%) and delays with projects and fellowships that had already been funded (30%).

> 'If I cannot submit a final data request before my grant expires, I will not have any funding to cover the costs of the data.'

The rest of this report focuses on those participants who had accessed data or were in the process of applying.

Consultation participants who had or were trying to apply for data

Over half had accessed data before (57% of 63), and a further quarter (25% of 63) had or were trying to: a total of 52 individuals. A quarter of participants had accessed or tried to access the 2014 Adult Psychiatric Morbidity Survey (APMS). A range of other survey datasets were also mentioned: earlier waves of APMS (14%), Health Survey for England (HSE, 19%), and the Millennium Cohort Study (19%). Three participants had tried or managed to access the 1999 or 2004 Mental Health of Children and Young People Survey (MHCYP), and nine (14%) mentioned the 2017 MHCYP (although no permissions had been granted at the time of writing (October 2020) for researchers to use this dataset). Participants were also interested in routinely collected datasets such as the National Pupil Dataset (NPD, 19%), Clinical Practice Research Datalink (CPRD, 15%), SAIL (4%), and the Mental Health Service Dataset (2%).

The median time taken for data access applications was 12 months. Following a discussion of consultation findings regarding the complexity of communication around data access processes, we discuss specific data access requirements that consultation participants found problematic. We go on to report the impacts of these problems, providing more detail about the amount of time consultation participants have spent navigating the data access process.

Complexity and communication within data access processes

By far the most common concern, cited by 41 (81%) consultation participants, was that the data access process was unclear and overcomplicated. While this issue was widely anticipated by those starting on the process, it was reported by nearly all participants who had managed to complete an application. This indicates that it is not just a lack of clarity about the process or a misplaced perception, but that the process itself is experienced as complicated and opaque:

> 'I have now found it very difficult to get permission (or work out how to get permission) in order to work directly with the…data. I basically gave up.'

Problems with communication were cited by 36 (69%) consultation participants. This included problems with communication between the data controller and the research team, and problems between the data controller and other agencies involved in the process. Compounding this, 15 consultation participants said they received inconsistent advice from different personnel working for the organisation holding the data.

Participants described how the data application process required involvement of university administration and senior staff, and that resourcing was sometimes not available for this.

> 'The fact that getting the [data] is impossible without contacting information management consultants and lawyers in our organisation despite it being anonymised is a sad waste of an important national resource.'

Seven participants (13%) said that university administration was unable or unwilling to put resource into supporting researchers' data applications. And 8 participants (15%) struggled to secure a sufficiently senior person to authorise final approval of applications, given the restrictions on who NHS Digital would accept (usually a formal Head of School or Division, who may have little capacity for engaging with individual applications for a dataset).

Specific data access requirements that were problematic

Registering with Data Access Request Service (DARS)
Half (50% of 52) of participants who were trying or had managed to get data had difficulties just registering with DARS (or its equivalent).

Perception of charge
Nearly half (44% of 52) perceived issues with charging. For routine datasets such as those mentioned above, where the costs of dataset curation is involved, there is indeed a charge levied. However, the majority of people applying for APMS data also cited concern about charges despite the fact that access to general population survey data, which is held at the UK Data Service, has always been free to data users. This perception doubtless stems from the fact that on top of each page of the online DARS application form it states that a fee of between £2,000 and £3,000 will be levied, although to date no fee is levied for access to many surveys.

Information Governance Toolkit/Data Security and Protection Toolkit
The NHS Digital Information Governance Toolkit has become a requirement for permission for data access to all NHSD datasets that NHS Digital controls, and is often awarded at institution-level, although it may be restricted to a department or even an individual. It changed names to the Data Security and Protection Toolkit in early 2019, however at the time of the consultation (May 2020), it was still referred to by its old name in the application system and so we refer to it thus here too (although the new name is now used in the online documentation). The application process for achieving this toolkit takes organisations a minimum of a year to achieve and involves substantial investment and staffing; the audited process currently includes over 300 requirements and accompanying proof. Many universities do not have this piece of NHS Digital documentation; 38% of the 52 participants applying or trying to apply for data stated that not having this form had caused problems for their application. In addition, a further 8% reported that while their institution had the toolkit, they

had struggled to get the details required for the application, such as the pass mark and date of renewal.

> 'One of the most onerous issues was providing lots of information governance assurances and details from our institution despite the fact that the data is anonymised and so in theory shouldn't need them but in practice they were demanded. This caused lots of difficulties because our information governance office kept replying "It's fine, this is not needed" when we couldn't move the application forward without it.'

Privacy notice

Despite the fact that survey data is provided in pseudonymised form, there is still a requirement for the researcher to post a 'privacy notice'[2]. The privacy notice has no real benefit for those taking part in survey research despite this being the intention of such notices. As researchers undertaking secondary analysis of such a dataset would not know the identity of study participants, they would be unable to act on any participants request for withdrawal. Some researchers feel compromised by being forced to post this publicly. More than a quarter of those we consulted (26%) raised issues with the privacy notice, either related to its production or to negotiations with institutions regarding where on their websites a privacy notice could be lodged.

Other specific requirements

Consultation participants highlighted a range of other specific requirements for an application that had been an issue for them. These included being able to:
- Demonstrate sufficient benefit to health services,
- Demonstrate sufficient impact,
- Produce required details of outputs,
- Produce acceptable data flow diagram,
- Produce acceptable study protocol.

The first two points above evoked surprise and frustration at the workshop run by MQ, given that much of the proposed research was perceived to have clear policy or practice implications. One participant described how they had abandoned one application for survey data as they were required to provide information it was not possible for them to know:

> '…there were lots of aspects of the application that seemed needlessly obstructive. For example, our study aimed to compare data from [a UK dataset] to datasets from other countries. We submitted the application and we were asked to say exactly which other countries' datasets we would be comparing it to. But obviously, this depends on the outcome of other applications and so it was impossible to say.'

Participants tended to cite more than one specific issue, indicating that individual requirements were not so much the problem as the sheer accumulation of multiple

requirements in forms that were not straightforward to meet and a process that was difficult to understand.

Impact on research

Time to complete an application

27 participants (52%) reported that they did not have the time alongside their other duties as researchers, scientists and academics to devote the time required by the application process.

> 'It's worth saying I've effectively abandoned my application for the [data]. I hope to return to it at some stage but…it became unfeasible as a project and so we have abandoned the research we were intending to do.'

> 'The impact on all my other work as it has all been so incredibly time consuming. It has taken weeks of my time…I have little to show for my efforts. The costs of this should be taken into account by data owners. They say they want to share data but their actions give entirely the opposite impression.'

Timing of projects and impact on current research

60% of those we consulted (n=31) had faced problems with project timelines due to greater than expected delays. They described how this had not only delayed projects, but had led to projects failing ever to start:

> 'Still awaiting permissions to link datasets two years on.'

> "The main issue I have experienced concerns timing. That is, both time taken for application to be approved and then more importantly from approval to data access. This can take up to two years."

> 'Project has still not been able to start, would have liked to have started six months ago.'

> 'We only received the data when the project funding had ended. We had a statistician but no data for 2.5 years, now we have data but no statistician or funding.'

> 'I have several projects, none of which I can currently progress.'

About three-quarters (73%) of the 52 consultation participants said that these delays had impacted on their ability to carry out research. Issues with current projects ranged in duration from 2 to 70 months (median 12 months, interquartile range (IQR) 6 months to 24 months, n=37). The majority of these people said that their specific issues remained unresolved and the application was ongoing, although some abandoned their data request:

> 'Gave up when realised access would be difficult/impossible (too risky for the timeline of the student).'

Figure 1. Length of time consultation participants reported they had been trying to gain access to the data, from the first contact with the data controller
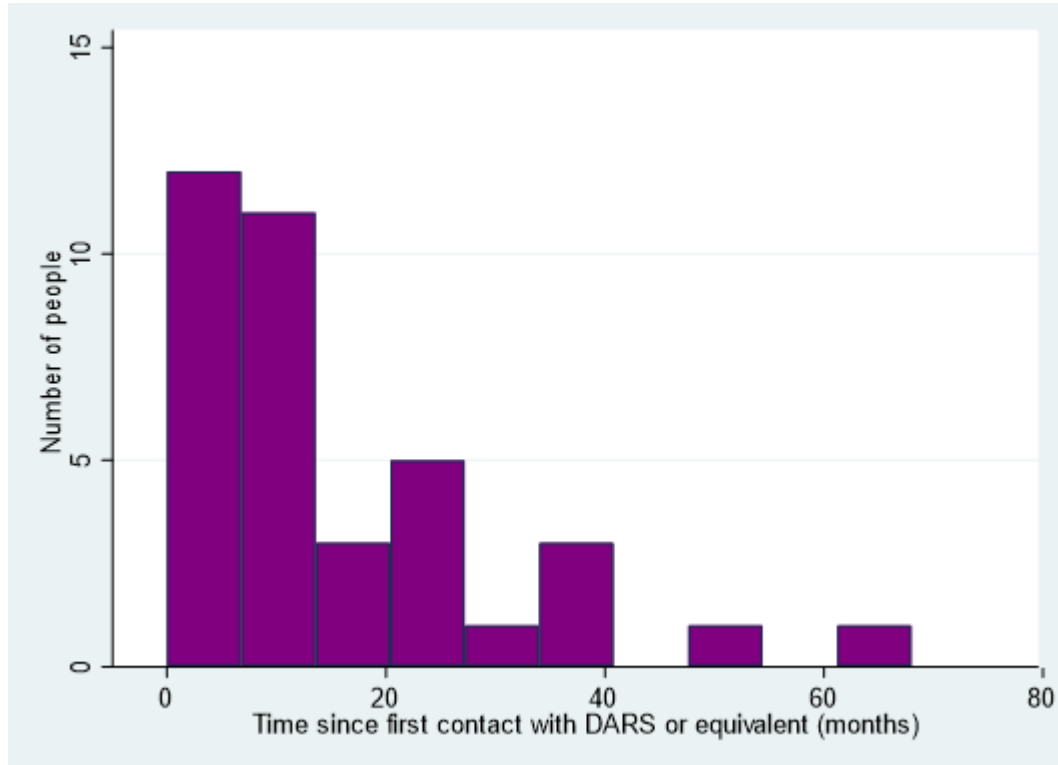
Where issues causing delays had been resolved, the median duration of delays was 6 months (IQR 4 months to 12 months, n=15), and where projects were still ongoing the median delay was 12 months (IQR 9 months to 15 months, n=12). The range of times reported is shown in Figures 2A and 2B. Where participants reported a range of values we used the minimum time given, so this is a conservative estimate of the delays experienced.

13

Figure 2A. Approximate length of delay where data access issues have now been resolved
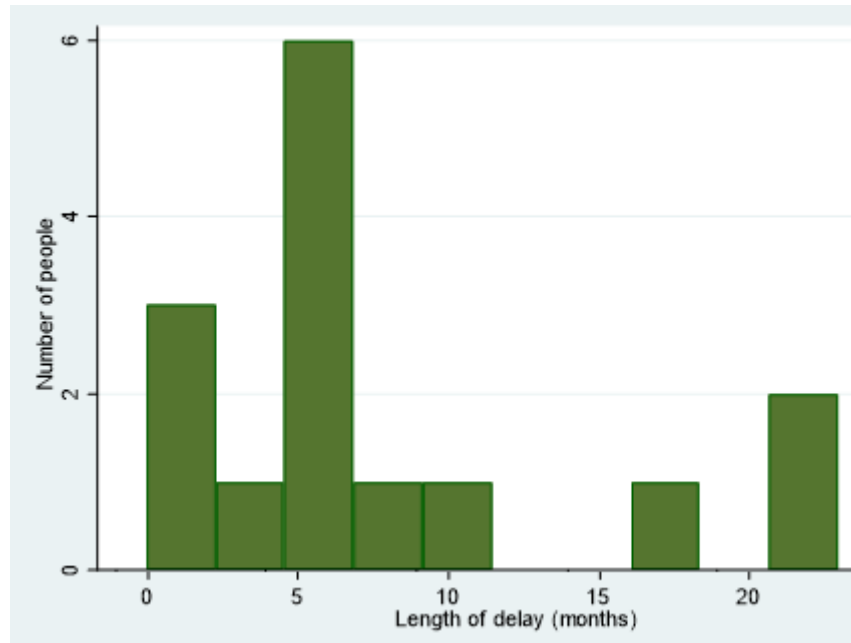
2B. Length of delay where data access issues remained unresolved
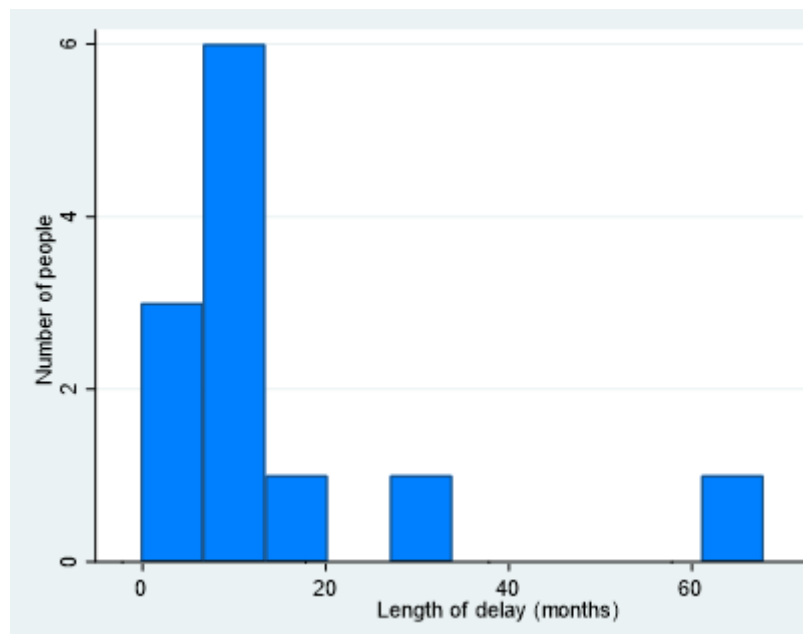
A quarter of participants (23%) said that they had encountered problems with funders because they could not demonstrate to them when applying for funding that they would be able to access data within a meaningful timeframe. Others described having grant funding withdrawn due to delays with data access, leading to unemployment.

Separate to the question on timelines, two-thirds (67%) of those we consulted said that difficulties with the data access process had impacted their ability to do research. This affected current projects, which had often been significantly delayed (16 consultation participants said that they got data so late there was insufficient time left for contracted staff to analyse it), resulting in altered or reduced scope (9 participants), or abandoned entirely (10 participants):

> 'Delay means dropped [survey] as source of data in a grant application'

Impact on future research
Consultation participants also described impacts on the nature and extent of future research. Many (35%) reported that because of their negative experiences, they were deterred from attempting to work these data sources in the future. A third (18 participants) said that barriers to access were resulted in postgraduate and undergraduate students not able to access important data sources.

> 'Was unlikely to get access to [data] for a student's dissertation despite [earlier round of data] being easily accessible from UK data service.'

Those we consulted reported that they are now choosing to avoid working with these health datasets entirely and trying to find alternative sources. One responded when asked if issues had been resolved, that they had 'given up'. Others replied:

> 'We have access to and have published studies on the 2000 and 2007 versions of this dataset and the contrast in how they are made available is really remarkable. 2000 and 2007 - they check you have a sensible research question and are a bona fide researcher and you are given access to the anonymised data. 2014 - **I have never dealt with such an onerous process for access to a dataset**.'

> 'The new arrangements…are at considerable risk of betraying the intention of the survey respondents to facilitate scientific research. Unless something is done to facilitate access to the data for purposes of analysis the value of the survey will be lost.'

Suggestions for improvement

We asked those we consulted whether they had any suggestions for improving data access procedures. Almost every consultation participant responded to this question. Suggestions fell into the broad categories of *improving communication, appropriate resourcing, unified and transparent procedures, improving access to easily anonymised data or smaller datasets*, and *increased trust* in academics. In addition, several consultation participants raised queries about the legality of some current data access restrictions and suggested external legal review is required. Specific suggestions to improve data access procedures within these broader categories are summarised below. However, underpinning suggestions for specific amendments to parts of the existing process was clear frustration and confusion with the process as a whole, highlighting a need for:

- Full review (to assess whether all demands were necessary and how to clarify and streamline the process) leading to
- A fundamental overhaul (to render the process more rational and proportionate).

Specific suggestions for improvement from which the recommendations arise:

Improved communication

- Coordinated communication channels between the research institution and data controllers (as opposed to individual researchers each having to attempt to navigate the data access process)
- Greater flexibility about who is authorised to sign-off data requests for institutions that is proportionate to the data requested and the organisation requesting it.
- Improved communication between data controllers and researchers, with a named individual (to permit cover for holidays etc), backed up with clear and up to date written information about the process and timeline.
- Data controllers to be sufficiently resourced to be able to have discussions with applicants and with information governance experts, and being able to meet demand in a timely manner.

Resources

- Accurate and transparent information around costs should be available from the outset of the application.
- Sufficient allocation of resource within the data controlling organisations to support data access processes, to ensure consistent communication from different staff members working in the same organisation, and to promote the efficient processing of applications.
- Investment in staff to facilitate data linkage requests.

<u>Procedures</u>
- Unified data access application procedures.
- More transparency of the process and likely timelines for negotiating with data controllers.
- Producing and agreeing data sharing agreements between involved parties .
- Case examples and completed templates from successful applications available to optimise the standard of submitted applications.

<u>Suggestions to improve access to appropriately 'anonymised' datasets</u>
- Recognising that data access should be proportionate to the type of data. Data from surveys and cohorts, which is provided with informed consent and can be pseudonymised should have fewer barriers to access than administrative data on people's personal experience with public sector services or data-linkage which may increase the risks of confidentiality breaches.
- Archiving anonymised data into the UK Data Service repository (reported by participants as easier to negotiate than other services); the UK Data Service could then negotiate a blanket access request covering bona fide academic researchers.
- Releasing or making available non-sensitive sections of wider datasets, even if there need to be restrictions due to safety concerns around some data.
- Curation of pre-linked datasets with data sharing agreements already in place for researchers to apply for specific variables or extracts to answer their research questions.
- Those with the knowledge of the data and linked data to set research questions that can be answered, and a process whereby researchers can apply to carry out the necessary analysis. This may be particularly facilitative for undergraduate and postgraduate research projects.

<u>Trust in academics</u>
- The default situation should be to treat researchers as professionals. It should be accepted that once a researcher has undergone safe data training, that they should be trusted to adhere to the legal agreements they sign. If they should try to identify participants or otherwise mismanage data, there is existing legal framework to manage this that can and should be applied.

<u>Legal issues</u>
- Some participants questioned whether strict data access procedures were in fact legal, where datasets were pseudonymised and patients or survey participants had provided full informed consent, with suggestions that external legal review was required, for example conducted by the Wellcome Trust:

'The [data] is a survey of the public and not of…patients. Participants consented to take part in good faith believing their time would be of value and made use of for the betterment of others. The Mental Capacity Act defines clearly the circumstances in which the wishes of an adult can be denied. All participants agreed to their data being used by bona fide researchers. Adults who did not appear capable of understanding and participating in the survey were not included. Any government organisation…denying access by bona fide researchers to [these] data is potentially over reaching its powers and if it refuses to desist should be challenged.'

Discussion and Recommendations

Of the 63 respondents to our consultation on data access, 52 had tried to or had accessed datasets and reported many challenges with the application process, while those who had not applied for data anticipated problems. Given the number of issues and complexity of processes reported by those we consulted who had tried to or had accessed data, these perceptions are not surprising. If action is not taken this may further undermine policy and practice related research. Our findings highlight that there are issues with data access procedures that mean that valuable datasets capturing a range of information are not being utilised as much as they could be. It is important to recognise that those who are contributing their individual data to research intend to contribute to scientific advancements and improvements to others' health and wellbeing, and we should therefore make efforts to ensure their contributions are not provided in vain. Access to these datasets could lead to important advances in preventing and treating health conditions, and understanding the impacts of lifestyle factors on health and wellbeing.

There are several concerning findings from the consultation:

Firstly, grant money and funding is being wasted, not awarded or even rescinded because of data access difficulties. Academic funding is always awarded based on the scientific contribution of the research alongside other factors and subject to peer review, while our consultation findings show that important work is not being completed because of administrative hold-ups and other overly onerous processes. Researchers are successfully convincing funders of the importance of their question to have data controllers reject the likely benefit or impact. Research projects are being entirely abandoned due to the scale of these difficulties, and as 'older' datasets from before 2014 are easier to access, some analyses are based on old data. This will undermine the usefulness of the findings to policy and practice about the epidemiology of health and mental health in the UK.

Since the advent of the GDPR in particular, the survey shows there have been unrealistic administrative barriers and increased bureaucracy. The GDPR has meant demands for documentation even when it should not be required due to anonymisation, and requests are being made that cannot be implemented because it is not logistically possible to do so with secondary data. A sign-off process that senior staff do not have time to properly engage with

renders the process meaningless. Privacy notices that cannot be implemented are a pointless requirement. In addition, while applicants can provide their research question and planned outputs, extensive details are dependent on results of research and dissemination process that may take place years into the future. Expectations that researchers can provide this are unrealistic and undermine the process of research if used as a reason to deny public benefit. Researchers also experienced inaccurate information and inconsistent advice, which would seem the easiest challenge to rectify.

GDPR was enforced for ethical reasons, to give tighter protection to people and their privacy. Instead, we argue its implementation has led, in the case of mental health data at least, to such large practical barriers to data access that many researchers simply cannot overcome them. This in addition, raises serious ethical concerns, particularly, that participants in research studies are under assurance that their responses will be used for the greater benefit of the wider community, yet the utility of the their information is greatly reduced by difficulties in the current system for access.

There is an obvious tension between the experiences of those we consulted and the open science movement(4). As many experimental scientists move to make data more easily accessible in the quest for transparency and rigour in scientific practice, it seems ironic that the potential of survey, cohort and administrative data to support evidence-based policy and practice is being undermined. Data on which findings are founded need to be are easy to access, so that others can replicate analyses directly and in alternative datasets.

All these additional barriers meant that participants in our consultation were less likely to access, or even to try to access data. Given that many other countries routinely publish studies using recent epidemiological data, this will clearly result in the UK lagging behind in the international scientific field. The importance of scientists and researchers being able to access recent epidemiological data cannot be understated. For several surveys, e.g. APMS and MHCYP, repeated waves of surveys allow scientists to track changes in population health over time. This allows them to identify and respond to concerning rises in e.g. self-harm, anxiety and depression in teenage girls, and to formally test whether changes in diagnosed disorders reflect true rises in prevalence or improved recognition of symptoms (e.g. Autism spectrum conditions). It appears from our survey that such scientists have to fall back on using data from over 10 years ago in order to complete their projects.

We also found that the length of time that it takes to resolve data access issues can stretch into years, and the median delay was 12 months (see Figure 1). Those without secure employment, such as junior academics or research assistants who work on short term contracts may be disproportionately affected by delays to data access. Data access issues therefore exacerbate pre-existing inequalities in career progression that are not only impacting on early career researchers, independent academics and analysts, but also on the next generation of researchers through difficulties accessing data for undergraduate and

postgraduate research projects. This is deterring the next generation of researchers from using large datasets and direct them away from policy and practice questions to the detriment of public mental health. Finally, it was clear from our consultation that data access systems need an overhaul. Half of our participants had difficulty simply registering with one service. This is the first step in what is described as a lengthy and difficult process, and participants reported difficulties with all stages of the process, including data not being provided in a timely manner when agreements are finally in place. This was considered to be unacceptable, and potentially illegal.

Consultation strengths and limitations

Given that this was a consultation, with respondents recruited via social media, organisational and personal networks, it may be that those who had experienced data access issues were more likely to complete the survey. This consultation was not intended to be representative, but rather to scope the range of issues that people were experiencing, and we consider 63 a good response rate from a small academic community. Regardless of the response rate across the community of potential respondents, the number of issues raised is alarming as is the consistency with which barriers were reported. We have captured information on a large range of problems with data access that quite a number of individuals are currently experiencing. These data provide clear signals about what would improve these processes going forward. The consultation focused on mental health data, but a recent open letter indicates similar frustration in relation to Child Health (3) and it would be interesting to assess whether the new legislation has had a similar impact in other fields of services and policy research.

Recommendations

Based on our consultation, we make the following recommendations for all stakeholders with a suggested time frame of 12 months:

All stakeholders, governmental level to lead:
- Redesign a proportionate and uniform process to balance the risks of breaches of privacy with the benefit to science and health policy based on the level of sensitivity of the data (e.g. low, medium, high).

Individual data-user level:
- Identify and undertake appropriate training relating to data security and GDPR.
- Demonstrate appropriate supervision from more experienced individuals where necessary.

Institutions hosting those requiring data access:
- Maintain infrastructure to support the necessary permissions, such as an Information Governance Toolkit/Data Security and Protection Toolkit.
- Ensure that appropriately trained senior staff have sufficient time to advise as well as to review clear timelines for the sign-off applications.

- Link those applying for data access with those who have successful applications to optimise initial applications.

Data custodians:
- Ensure transparent, consistent and clear information about the access process with detailed information about what is required at each step.
- Provide additional low-risk open access datasets that can be shared with institutions for student projects to ensure that we encourage and develop the next generation of data scientists.
- Link those applying for data access with those who have successful applications to optimise initial applications.

Governmental level
- Establish an all-party parliamentary group to review how to optimise safe and legal access to data.
- Clarify the remit of data controllers, data curators and data processors.
- Identify and share examples of best practice by data controllers and organisations that facilitate the sharing of data (e.g the ESRC UK Data Service).

Conclusion

In conclusion, those seeking to access large health-related datasets in the UK report expecting and experiencing a huge array of difficulties, undermining research that aims to support policy and practice. Many of these issues occurred because of poor communication, lack of appropriate resources and overly complex processes and procedures. We have made several recommendations for improving access procedures that data controllers and organisations hosting individuals wishing to use such data could enact. Given the consistent complaints of lack of clarity and poor communication, it is not surprising if researcher error contributes to the delay, but the remedy to this is easy and clear; improved clarity and communication about the process. Researchers are keen to work with other stakeholders to make this process run more smoothly as it is all our interests to do so.

It is clear that the current processes can take a prohibitively long time to navigate, and the impacts of this are widespread, including the failure to complete important studies, adverse impacts on individual careers or employment particularly for early-career researchers, and financial impacts on individual projects, funders and hosting organisations. In addition to this, an inability to conduct analyses using contemporary health data will detrimentally impact on the health of the nation and risks the UK's scientific advances falling behind the international field. The UK simply cannot expect to shine while it forces researchers to use outdated information. Tying a muddle of misguided red tape around its most recent and precious data gifted by participants precisely because they want to help research perverts their generosity and leaves the nation disadvantaged.

Glossary

[1] Pseudonymised means replacing or removing any information that someone could use to identify the participant (accidentally or otherwise), for example place names.

[2] A privacy notice, according to GDPR, is a public document from an organisation that explains how that organization processes personal data and how it applies data protection principles. For more information on GDPR and privacy notices, see https://gdpr.eu/privacy-notice/

Appendix

Survey introduction

We are interested to hear from all who are interested or involved in the quantitative analysis of survey or administrative data, including students, practitioners, commissioners, policy makers and those working in the third sector.

We are a group of experienced researchers who regularly work with such data and are working with MQ to improve processes around access. We want to learn from your experience and the answers you provide will be collated with all those we receive to produce a report that will support discussions with data controllers about how to facilitate access to data for those who need it. This report may be published, and any data included will be anonymised.

Please take the time to answer this brief, anonymous survey; we do not ask for any identifying details. Your continuing on to the survey questions implies your consent to our analysis of the data you provide.

References

1.   Office for Statistics Regulation. NHS Digital data sharing and access review: update. April 2020. https://osr.statisticsauthority.gov.uk/publication/nhs-digital-data-access-update/
2.   The Lancet Psychiatry. Smorgasbord or Smaug's hoard? Editorial, The Lancet Psychiatry. 2019;6(8):631.
3.   Cavallaro F, Robling, M., Lugg-Widger, F., Cannings-John, R. et al. Open letter to the ICO, CMOs and UK data providers: Reducing barriers to data access for research in the public interest–lessons from COVID-19. The BMJ (opinion). July 2020
4.   Allen C, Mehler DM. Open science challenges, benefits and tips in early career and beyond. PLoS biology. 2019;17(5):e3000246.