The challenges and opportunities of mental health data sharing in the UK

Ford, T., Mansfield, K. L., Markham, S., McManus, S., John, A., O'Reilly, D., Newlove-Delgado, T., Iveson, M., Fazel, M., Das Munshi, J., Dutta, R., Leavey , G., Downs, J., Foley, T., Russell, A., Maguire, A., Moon, G., Kirkham, E. J., Finning, K., ... Shenow, S. (2021). The challenges and opportunities of mental health data sharing in the UK. *The Lancet Digital Health*, *3*(6), e333-e336. https://doi.org/10.1016/S2589-7500(21)00078-9

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

The challenges and opportunities of mental health data sharing in the UK

The UK's National Health Service (NHS) generates uniquely rich data that should be rapidly deployed for policy and service improvement, yet researchers report difficulties in accessing these data. Paradoxically, these restrictions are occurring at the same time as the open science movement, which encourages data sharing to improve the rigour, transparency, and replicability of research. We describe the urgency of improvements to data access and propose solutions from a mental health research perspective, although the issues discussed extend to all areas in which analysis and linkage of health data support policy and practice. Actions are needed at every level, from data users and data custodians to government (panel).

Administrative datasets permit researchers to explore real world experiences, antecedents, and outcomes. Furthermore, record linkage to administrative data enhances the utility of research datasets and supports study of mental health over the life course.[1] NHS England provide aggregated statistics on their website, but accessing patient-level data, necessary for epidemiological data analysis, requires approved researcher status and application through the Data Access Request Service, a process that researchers are reporting to be prone to increasing delays.

Processes to access data have become longer and more complex, often requiring considerable perseverance, which we describe in detail in our survey (appendix pp 9–27). Despite overarching European and UK legislation, access also differs considerably between data custodians, with substantial variations in the application of legislation between the UK nations (appendix pp 2–8).

A major consideration is whether the data requested are identifiable. Since May, 2018, the General Data Protection Regulation (GDPR) has provided a strict legal framework for data protection across Europe; health data are considered special category data (highly sensitive). The GDPR sets out lawful bases for processing identifiable data; scientific research does not always need to rely on consent and is potentially exempt from one or more obligations set out by the GDPR, considered on a case-by-case basis when certain criteria can be documented.

Organisations and individuals have needed to adjust to the new legal framework and guidelines around the definition of personal data, which explicitly includes pseudonymised data. Currently, the border dividing personal (including pseudonymised) from deidentified (anonymised) data is unclear (appendix p 1), and depends on the estimated likelihood that access could lead to the identification of individuals.[2] Data that were previously considered to be anonymised are now being scrutinised again under the new definitions. Given the novel legislation and worrying data breaches,

---

***Panel:* Actions needed over the next 12 months**

All stakeholders, led at the governmental level
- Redesign a proportionate and uniform data access process that balances the risks of privacy breaches with the benefits to science and health policy using the data sensitivity level (low, medium, or high)

Individual data user level
- Identify and undertake data security and General Data Protection Regulation training
- Obtain appropriate supervision from experienced individuals where necessary
- Maintain regular contact with data custodians, providing updates on changing requirements and project circumstances

Institutions hosting data users
- Maintain infrastructure to support the necessary permissions, such as obtaining and maintaining a National Health Service Data Security and Protection Toolkit
- Ensure that appropriately trained senior staff have sufficient time to advise, review, and sign off applications
- Link those applying for data access with those who have had successful applications

Data custodians
- Ensure that there is transparent, consistent, and clear information about the access process and 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
- Maintain regular contact with project leaders, providing updates on data availability and application progress
- Link those applying for data access with those who have had successful applications

Governmental level
- Establish an All-Party Parliamentary Group to review how to optimise the safe and legal access to data, including a review of how legal requirements are being interpreted
- 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 (eg, the UK Data Service)

---

there is understandably caution around sharing even deidentified data.

The Information Commissioner's Office perceive the GDPR as facilitative of research, but many researchers report otherwise. Under the GDPR, special category data can only be processed under specific conditions. When this processing is for research purposes, one condition is to "demonstrate that the processing is in the public interest", which can be difficult for research that has important but delayed benefits, such as exploratory analyses. Precedent in terms of previous research has lost its power to persuade data custodians, and many researchers struggle to produce evidence of rapid benefit to the NHS or the public.

Delayed access to data is undermining research activity and deterring the next generation of data scientists. Delays have increased over the past decade, but were hugely amplified after the introduction of the GDPR, and remain considerable despite concerted efforts (appendix p 20). Changing application processes are complicated by changes in the names of datasets and data controllers. Useful information sources have ceased to exist, while inconsistent information between agencies causes confusion. The resource implications for government departments experiencing chronic staff shortages and rapid staff turnover must be recognised and remedied. Delays in data access reduce researchers' trust in data controllers, discouraging future applications.[3]

In our survey (appendix pp 9–27), 36 (69%) of 52 researchers reported communication difficulties during the application process and 38 (73%) reported delays in accessing data. Analyses of the Adult Psychiatric Morbidity Surveys suggest a substantial reduction in applications, downloads, and papers in recent years.[4] By 2011, the 2007 survey dataset had been downloaded 165 times, leading to more than 30 papers, whereas just two papers were published within 4 years of the 2014 survey. Since 2016, the 2007 survey dataset has been downloaded by a further 288 data users, and the 2014 dataset by a further 22.

The impact of improving access to administrative datasets would be profound. There would be more evidence generated for policy makers, commissioners, and practitioners. Research resources could be deployed towards analysis instead of being diverted to the application process. There would be a reduction in the number of projects that are limited or abandoned as grants and budgets expire (appendix pp 9–27). Research could be done on up-to-date data, rather than by relying on older datasets that are often easier to access, and findings would be more relevant to current public health priorities.

Rapid access to relevant data is especially important to understand and mitigate the negative effects of the COVID-19 pandemic.[5] There is a growing need for data science skills, yet students and early career researchers often do not have time to spend on applying for data, diminishing recruitment and the training of individuals in such skills. Many datasets and research studies are publicly funded, including the time spent completing and scrutinising applications for access, and this funding could be used more effectively.

Faster access requires less bureaucracy. Data custodians and researchers are collaborating to transform application processes while adhering to current legislation. The COVID-19 pandemic has shown the extent and speed of change that is possible, but permanent solutions are required, including smoother collaboration between organisations that control the data.[6] A fundamental cross-institution commitment to data sharing and onward linkage is essential, and requires transparency by data controllers to data providers and researchers around the potential linkage of personal data and safe sharing for research. Codesign of the data access process by institutions would ensure that governance and access processes are clear and consistent. Ideally, technical and legal solutions would be nationally agreed and implemented by government organisations and data controllers.

Public and professional trust in the system is essential to the entire process. To build this trust, the UK Government, plus social and health-care organisations responsible for sharing and using data, need to show robust and secure processes, and clearly communicate the benefits of sharing data with researchers. Data collection and sharing needs to be embedded in routine service provision by the NHS and controllers of health data, and supported by improved digital information technology, enabling the collection of better data, leading to greater impact.

Complexity might be reduced if circumstances could be defined, by legal and governance experts working with the controllers of large-scale administrative datasets,

in which pseudonymised data can be considered anonymous.[2] Reidentification becomes more probable when data are combined or seen in a new context, or if uncommon responses are present. Some people question whether large-scale individual-level data could ever be fully anonymous and suggest using synthetic data, which are not directly measured but artificially constructed to mimic directly measured data.[7] Epidemiological research, however, requires individual-level data in which the dynamic relationships between factors are intact. Rather than waiting for methods by which individual-level data can be fully anonymised, data controllers and researchers must rely on sound information governance and contracts that ensure the data will not be reported or shared in ways that risk identifying individuals.

The application process should be proportionate to the anonymity and sensitivity of the data (appendix pp 21–27), concepts that need to be understood by researchers and data custodians. Accredited researcher training ensures that researchers understand data governance. For more sensitive data, centralised data hubs that assist data linkage, sourcing, and curation, and secure data environments for analysis, with the vetting of outputs, could ensure confidentiality. Efficiency will depend on the level of staffing versus demand, which should be monitored. Streamlining the data access process to focus on the strictly necessary would reduce the burden on both data custodians and researchers. There is increasing recognition that the creation of one-off, large, multipurpose research datasets might be an efficient way to meet the needs of the research community.[8]

Good communication between data custodians and researchers is essential. Researchers must be responsive to questions raised by reviewers, and data custodians should keep researchers informed of progress or delays. Having consistent named contacts in both the research institution and the institution holding the data would reduce inappropriate submissions. A research buddy system, supported by those familiar with the data and metadata, could improve projects and facilitate a robust data access process, ensuring that knowledge is not lost when key staff move on.

The UK Government's Life Sciences Strategy aims to position the UK as a global leader in data. Impediments to accessing data stand to derail this goal. A new Government review aims to improve both the efficiency and safety of health data access for research. The operating framework for researchers to access data must become transparent, uniform, responsive, and consistent. The advancement of knowledge for public benefit requires timely access to data for those qualified to use it responsibly, to avoid potential harm due to out-of-date information.[9] Researchers build on each other's conclusions to work towards overall outcomes that are greater than the sum of their individual efforts. Only with different researchers applying their scientific rigour to the same data can the accuracy, analyses, and conclusions be verified. Making data readily accessible, without undue delays, is key to this process; after all, "the value of data lies in their use".[10]

Tamsin Ford, *Karen L Mansfield, Sarah Markham, Sally McManus, Ann John, Dermot O'Reilly, Tamsin Newlove-Delgado, Matthew H Iveson, Mina Fazel, Jayati Das Munshi, Rina Dutta, Gerard Leavy, Johnny Downs, Tom Foley, Abigail Russell, Aideen Maguire, Graham Moon, Elizabeth J Kirkham, Katie Finning, Ginny Russell, Anna Moore, Peter B Jones, Sarah Shenow
karen.mansfield@psych.ox.ac.uk

Department of Psychiatry, University of Cambridge, Cambridge, UK (TFor, AMo, PBJ); Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK (KLM, MF); Department of Biostatistics and Health

For more on **accredited researcher training** see https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme

For the **new Government review** see https://www.gov.uk/government/news/new-review-into-use-of-health-data-for-research-and-analysis

Informatics (SMa) and Department of Psychological Medicine (JDM), Institute of Psychiatry, Psychology and Neuroscience (JD), and Department of Psychological Medicine, Division of Academic Psychiatry (RD), King's College London, London, UK; School of Health Sciences, City, University of London, London, UK (SMc); Public Health and Psychiatry, Swansea University Medical School, Swansea University, Swansea, UK (AJ); Centre for Public Health, Queen's University Belfast, Royal Group of Hospitals, Belfast, UK (DO, AMa); Children and Young People's Mental Health Research Collaboration, University of Exeter College of Medicine and Health, University of Exeter Medical School, Exeter, UK (TN-D, AR, KF, GR); Centre for Clinical Brain Sciences (EJK) and Centre for Cognitive Ageing and Cognitive Epidemiology (MHI), Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, UK; School of Psychology, Ulster University, Londonderry, UK (GL); National Institute for Health Research Maudsley Biomedical Research Centre, South London and Maudsley National Health Service Foundation Trust, London, UK (JD); Population Health Sciences Institute, Newcastle University, Newcastle, UK (TFol); School of Geography and Environmental Science, University of Southampton, Southampton, UK (GM); MQ, London, UK (SS)

1 Rahman MA, Todd C, John A, et al. School achievement as a predictor of depression and self-harm in adolescence: linked education and health record study. *Br J Psychiatry* 2018; **212:** 215–21.

2 Mourby M, Mackey E, Elliot M, et al. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. *Comput Law Secur Rev* 2018; **34:** 222–33.

3 Sexton A, Shepherd E, Duke-Williams O, Eveleigh A. A balance of trust in the use of government administrative data. *Arch Sci* 2017; **17:** 305–30.

4 The Lancet Psychiatry. Smorgasbord or Smaug's hoard? *Lancet Psychiatry* 2019; **6:** 631.

5 Holmes EA, O'Connor RC, Perry VH, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *Lancet Psychiatry* 2020; **7:** 547–60.

6 Harron KL, Doidge JC, Knight HE, et al. A guide to evaluating linkage quality for the analysis of linked data. *Int J Epidemiol* 2017; **46:** 1699–710.

7 Rocher L, Hendrickx JM, de Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* 2019; **10:** 3069.

8 Davies M, Jones H, Conolly A. Public attitudes to data linkage. March, 2018. https://natcen.ac.uk/our-research/research/public-attitudes-to-data-linkage/ (accessed April 26, 2021).

9 Jones KH, Laurie G, Stevens L, Dobbs C, Ford DV, Lea N. The other side of the coin: harm due to the non-use of health-related data. *Int J Med Inform* 2017; **97:** 43–51.

10 National Research Council. Bits of power: issues in global access to scientific data. Washington, DC: The National Academies Press, 1997.