Check for updates

CASE STUDY

Data sharing in child and adolescent psychiatry research: Key challenges (and some potential solutions)

[version 1; peer review: 1 approved]

Beth Oakley[1], Alexandra Lautarescu[1,2], Tony Charman[2,3], Christopher Chatham[4], Eva Loth[1], Christian Beckmann[5,6], Thomas Bourgeron[7], Florence Campana[7], Rosie Holt[8], Eliza Eaton[8], Pierre Violland[9], Katrien Van den Bosch[9], Siofra Heraty[10], Scott Wagers[11], Jan K. Buitelaar[5,12], Declan Murphy[1,3,13], Amy Goodwin[1], Emily Jones[10,14,15], The AIMS LEAP Group

[1]Department of Forensic and Neurodevelopmental Sciences, King's College London Institute of Psychiatry Psychology & Neuroscience, London, England, UK
[2]Department of Psychology, King's College London Institute of Psychiatry Psychology & Neuroscience, London, England, UK
[3]South London and Maudsley NHS Foundation Trust, London, England, UK
[4]F. Hoffman-La Roche Ltd, Basel, Switzerland
[5]Department of Cognitive Neuroscience, Radboud University Donders Institute for Brain Cognition and Behaviour, Nijmegen, Gelderland, The Netherlands
[6]Centre for Functional MRI of the Brain, University of Oxford, Oxford, England, UK
[7]Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, Île-de-France, France
[8]Autism Research Centre, University of Cambridge Department of Psychiatry, Cambridge, England, UK
[9]AIMS-2-TRIALS Autism Representative (A-Rep), Cambridge, UK
[10]Birkbeck University of London Centre for Brain and Cognitive Development, London, England, UK
[11]BioSci Consulting, Maasmechelen, Belgium
[12]Karakter Child and Adolescent Psychiatry University Center, Nijmegen, The Netherlands
[13]Institute for Translational Neurodevelopment, King's College London Institute of Psychiatry Psychology & Neuroscience, London, England, UK
[14]Department of Child and Adolescent Psychiatry, King's College London Institute of Psychiatry Psychology & Neuroscience, London, England, UK
[15]MRC Centre for Neurodevelopmental Disorders, MRC Centre for Neurodevelopmental Disorders, London, England, UK

v1 **First published:** 31 Mar 2025, **5**:93
https://doi.org/10.12688/openreseurope.18147.1
Latest published: 31 Mar 2025, **5**:93
https://doi.org/10.12688/openreseurope.18147.1

Abstract

Background

The field of biomedical research is entering a new era, in which public data sharing is increasingly the norm. There are many advantages of embracing data sharing initiatives, including tackling the replication crisis through enhanced transparency and publication of null findings, facilitating global collaborations to accelerate research progress, enhancing cost-effectiveness by reducing duplication of efforts, and

Open Peer Review

Approval Status ✓

	1
version 1 31 Mar 2025	✓ view

1. **Elizabeth Kirkham**, University of Edinburgh, Edinburgh, UK

Any reports and responses or comments on the article can be found at the end of the article.

making scientific advances more accessible to the public.

However, there are also several crucial ethical and logistical challenges that must be addressed to maximise the benefits of data sharing and minimise risks. The potential, and increasingly recognised, risks of unregulated data sharing (e.g., data reidentification, misuse, and lack of representativeness due to variability in who agrees to share data) have also been exemplified by high profile data breaches and directly clash with efforts to make research more robust, accessible, and global.

Methods/Results

Here, we narratively outline current challenges for data sharing from the perspective of child and adolescent psychiatry, one area where they may be particularly acute. For example, child and early adolescent research often requires caregivers to consent on behalf of a minor – increasing the responsibility of researchers to consider how the science of today may evolve into the future (when those individuals are no longer minors). We use data from our research consortium Autism Innovative Medicines Study - 2 - Trials (AIMS-2-TRIALS; https://www.aims-2-trials.eu/) to illustrate the points raised in this perspective piece.

Conclusions

We also propose some potential solutions to begin to address current challenges for data sharing, focusing on key priorities, including shared control of data curation between researcher and participant communities and equity of access by research groups to the tools and resources needed to conduct responsible and sustainable data sharing.

Keywords
Data, data sharing, child and adolescent psychiatry, developmental psychiatry, global research, ethics.

H2020 This article is included in the Horizon 2020 gateway.

Corresponding author: Beth Oakley (bethany.f.oakley@kcl.ac.uk)

Author roles: Oakley B: Conceptualization, Data Curation, Formal Analysis, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Lautarescu A**: Conceptualization, Writing – Review & Editing; **Charman T**: Conceptualization, Data Curation, Funding Acquisition, Writing – Review & Editing; **Chatham C**: Conceptualization, Writing – Review & Editing; **Loth E**: Data Curation, Funding Acquisition, Project Administration, Writing – Review & Editing; **Beckmann C**: Conceptualization, Data Curation, Funding Acquisition, Writing – Review & Editing; **Bourgeron T**: Funding Acquisition, Writing – Review & Editing; **Campana F**: Data Curation, Writing – Review & Editing; **Holt R**: Conceptualization, Writing – Review & Editing; **Eaton E**: Conceptualization, Writing – Review & Editing; **Violland P**: Conceptualization, Writing – Review & Editing; **Van den Bosch K**: Conceptualization, Writing – Review & Editing; **Heraty S**: Writing – Review & Editing; **Wagers S**: Conceptualization, Writing – Review & Editing; **Buitelaar JK**: Data Curation, Funding Acquisition, Methodology, Writing – Review & Editing; **Murphy D**: Data Curation, Funding Acquisition, Supervision, Writing – Review & Editing; **Goodwin A**: Project Administration, Writing – Review & Editing; **Jones E**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: In the past 3 years, TC has served as a paid consultant to F. Hoffmann-La Roche Ltd. and Servier and has received royalties from Sage Publications and Guilford Publications. DM has received funding for a PhD studentship from Compass, and for consulting from Jaguar Therapeutics and Hoffman Le Roche. CC was a full-time employee of Genentech and owned stocks or RSUs in Roche Holdings, Ltd at the time of preparation of this manuscript. JKB has been a consultant to, advisory board member of, and a speaker for Takeda/Shire, Medice, Roche, and Servier. He is not an employee of any of these companies and not a stock shareholder of any of these companies. He has no other financial or material support, including expert testimony, patents, or royalties. CFB is director and shareholder in SBGneuro Ltd.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No [777394] Autism Innovative Medicine Studies – 2 – Trials [AIMS-2-TRIALS].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Oakley B *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License , which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Oakley B, Lautarescu A, Charman T *et al.* **Data sharing in child and adolescent psychiatry research: Key challenges (and some potential solutions) [version 1; peer review: 1 approved]** Open Research Europe 2025, **5**:93 https://doi.org/10.12688/openreseurope.18147.1

First published: 31 Mar 2025, **5**:93 https://doi.org/10.12688/openreseurope.18147.1

Introduction

Background

Translating research progress to public health impact remains a major challenge for child and adolescent psychiatry (Insel, 2009). Despite the substantial heterogeneity of mental health conditions, interventions are still predominantly targeted by diagnostic group, and the development and implementation of novel treatments has been slow (Dalgleish *et al.*, 2020; Stringaris, 2014). Consequently, while rates of psychiatric intervention have substantially increased, less than half of individuals experiencing clinically relevant mental health problems receive an evidence-based intervention and fewer still respond positively (Insel, 2009).

To respond to this challenge, big data in psychiatry has been a key area of progress (Ressler & Williams, 2021; Stein *et al.*, 2022) – the bringing together of large datasets (via key initiatives, such as ELIXIR; https://elixir-europe.org/) that allow for the mapping of candidate transdiagnostic mechanisms that underpin specific symptoms and represent targets for individualised intervention.

The emerging data sharing mandate

Despite the potential promise of big data in psychiatry, the sharing of data between research groups and in the public domain to facilitate big data analytics approaches has long posed both ethical and logistical challenges, which are increasingly coming to the fore (Ford *et al.*, 2021a; Monteith *et al.*, 2015; Mostert *et al.*, 2018a). This is, perhaps, particularly poignant in the field of psychiatry (and life/biomedical sciences more broadly), where research data captures human characteristics and experiences and often includes sensitive information, such as demographic factors (e.g., ethnicity, sex, gender), mental and physical health, qualitative data on individual experiences and attitudes (Kirilova & Karcher, 2017), and/or neuroimaging and genetic data (Ford *et al.*, 2021b). An added consideration in child and adolescent psychiatry is that key decisions regarding research participation and data use may be made by a legal guardian (e.g., parent) on behalf of a young person (Patrinos *et al.*, 2022a). Thus, when the young person becomes legally able to consent for themselves, it is possible that their own preferences may differ from those earlier indicated by their legal guardian, and particularly if the research landscape (i.e., the purpose and impact of the research to which data were contributed) has significantly evolved in this time.

Given these (and many other) complexities, ethics review boards have previously tended to require researchers to provide assurances that data will only be accessible by a restricted local research team who collect, process, analyse, and report on such data (Meyer, 2018). More recently, however, there has been a shift toward promoting open research principles that aim to improve the transparency, efficiency, and quality of research, including through data sharing (Munafo *et al.*, 2017; Towse *et al.*, 2019). In fact, journal publication and funder policies are increasingly making data sharing a *requirement*. For instance, the US National Institutes of Health (NIH) published an announcement in 2022 that they would mandate public data sharing for successful grant applicants (see Kozlov, 2022). Furthermore, regardless of dissemination and funding requirements, data sharing between research groups is becoming increasingly integral to scientific progress, as a substantial number of research projects (including the authors' own consortia) now rely on the sharing of expertise and resources across multiple collaborating sites around the world (Kalkman *et al.*, 2019a).

As indicated above, there are many advantages of embracing data sharing. Through facilitating global collaborations to accelerate scientific progress, data sharing can contribute to improving health and wellbeing outcomes for individuals and families, enhancing efficiency and cost-effectiveness, promoting collaboration, and minimising participant burden by reducing duplication of efforts. Data sharing can make scientific advances more accessible to the public, enhance trust in science, and contribute to tackling the replication crisis through enhanced transparency and accountability, better data aggregation and management practices, and facilitation of replication and data reuse (Chan *et al.*, 2014; Levenstein & Lyle, 2018; Tenopir *et al.*, 2011).

Nevertheless, the potential risks of data sharing have come under increasing scrutiny in recent years - partly as a result of major data breaches (e.g., the Facebook–Cambridge Analytica data scandal; Schneble *et al.*, 2018) and concerns of data misuse (e.g., the commercialisation of personal data; Howe *et al.*, 2018a; Watson *et al.*, 2022a), both implicated in the onset of new legislative frameworks to tackle these risks (e.g., General Data Protection Regulation, or GDPR, in Europe/UK; https://gdpr-info.eu/).

Hence, there exists some (often unacknowledged) tension between the goals of data sharing and the ethical, legal, and logistical frameworks within which these goals can and should operate. In this paper, we outline some of the current challenges for data sharing from one area where they may be particularly acute - the field of child and adolescent psychiatry. Moreover, we propose potential solutions to begin to address these challenges, focusing on key priorities of shared data curation (between participant and researcher communities), supportive infrastructure (including the funding landscape and legal/ethical frameworks for research), technical solutions, and public engagement.

Challenges for data sharing in paediatric research

Ethical considerations

Ethical considerations for data sharing in paediatric science can be broadly conceptualised in relation to the four pillars of medical ethics: beneficence (to do good), non-maleficence (to do no harm), autonomy (individual choice), and justice (promoting fairness; Ross *et al.*, 2018). Data sharing has clear potential for beneficence. It can enhance the transparency of research for both the scientific community and the public, and expedite important scientific breakthroughs, including those with direct impact on communities - for example, those that address a community priority, improve early identification of mental health difficulties, improve support strategies, or influence health and

social care policy (Walport & Brest, 2011). Nevertheless, ensuring non-maleficence, autonomy, and justice is contingent on the data sharing practices implemented. The challenge here is that, when working with human data, there is no 'one size fits all' method for removing (or, more often, substantially minimising) the potential risks associated with data sharing, such as participant reidentification or data misuse.

Avoiding participant reidentification. A key example of this challenge is centred on the level of anonymisation of datasets to be shared. Currently, open science predominantly endorses the sharing of fully anonymised datasets - that is, data that has been completely stripped of all potentially identifying information, also including removal of research participant codes/pseudonyms (e.g., 'Participant 101'). In principle, it is perceived to be relatively simple for research teams to anonymise datasets in this way. However, in practice, the increasingly sophisticated multimodal research methods now commonplace in paediatric science mean that the combination of information about an individual drawn from across multiple data types increases the possibility of (inadvertent) participant reidentification (Mostert *et al.*, 2018b; Nature Biotechnology, 2023). This may be particularly true for research including particular data types, such as biometric data and genetic sequences (especially when applying multiomics approaches) that embed complex information about an individual, and/or where data types become more individuating as science progresses (e.g., fMRI fingerprinting; Finn *et al.*, 2015). In addition, even minimal demographic information (e.g., maternal age combined with child age and sex at birth) may be identifying in some paediatric populations, such as children born with very rare genetic conditions or extremely preterm - though these groups have also been found to have relatively positive perceptions of data sharing, linked to a motivation to support other families (please also see Begum Ali *et al.*, 2023).

Other data types of high value to paediatric science are inherently 'sensitive' (as defined by GDPR), or even directly identifiable. These include, for instance, information about ethnicity, religion, and physical and mental health, as well as videotapes of research assessments with children that require later coding (e.g., Bayley Scales of Infant and Toddler Development, Bayley, 2005; Parent-Child Interaction tasks). Although directly identifiable materials are very rarely shared publicly, and technologies have become increasingly sophisticated to remove or reduce identifiable information embedded within data (e.g., defacing MRI scans), specified data transfers between global research partners within a single overarching scientific project may be necessary to access and share expertise (e.g., to perform data quality procedures, or specialist analytics). Furthermore, it is possible that some data types that are not currently categorised as 'sensitive' may become so in the future, long after consent preferences for sharing of that data have been ascertained, as societal norms and technologies (e.g., AI) change over time.

Overall, as demonstrated by the examples outlined here, it is frequently the most 'cutting edge' and valuable psychiatric research that is also the most individuating - for instance, where data on variation in early developmental trajectories across (say) behaviour, brain, and genetics is combined to make more informed recommendations about the optimal support strategies for a given individual (Hahn *et al.*, 2017).

Preventing data misuse. Equally important to the way that research data are shared (e.g., anonymisation) is the purpose of sharing. Indeed, even where datasets are fully deidentified, participants may still have concerns about what their data may be used for, and by whom. A high-profile illustration of this is the case of the Havasupai people who consented for researchers at Arizona State University to collect blood samples to investigate high rates of diabetes in this population, but later discovered that these samples had also been used for additional research that they did not consent to and was considered harmful to the cultural heritage of the community (Mello & Wolf, 2010). Other key examples represented in both the scientific literature and popular media include concerns over the potential downstream uses of genetic data (e.g., prenatal testing, commercialisation; Hoge & Appelbaum, 2012; Palk *et al.*, 2019), security concerns over facial recognition and other digital technologies (Gooding, 2019; Pavarini *et al.*, 2022), and possible implications for exposure to or reinforcement of stigma or discrimination in daily life if health information were openly accessible (e.g., employability, health insurance; Howe *et al.*, 2018b; Watson *et al.*, 2022b). Therefore, the purpose of sharing research data deserves considerable attention in its own right and to ensure non-maleficence - what if publicly shared data is used in ways that are not desired by, or are harmful to, the participant communities contributing that data?

Participant informed consent is a fundamental consideration here to enhance autonomy. It may not always be possible for research groups to be completely specific about every component of the projects and analyses that research data may be used for well into the future at the point where consent for data sharing is first sought. For example, where participants are contributing to a data resource, they may be informed of the broad purpose of use of their information (e.g., 'for researchers undertaking paediatric health research in the public interest'). However, there are also a range of unknowns (Ferretti *et al.*, 2021a; Garrison *et al.*, 2015; Jao *et al.*, 2015; Miranda Montoya *et al.*, 2022), including, which academic/non-academic individuals/groups may apply to access the data, and what their specific analyses will focus on in relation to the wider dataset available (e.g., epilepsy vs. anxiety). In addition, as noted earlier, one issue for informed consent is that an individual's preferences over the use of their research data may change over time and this is particularly relevant to paediatric samples where, due to age and/or capacity, the autonomy of young people over the use of their data may be reduced where a parent consented on their behalf (see also Hunter & Pierscionek, 2007). Differences between a parent and young person's consent preferences have been reported, with some evidence to suggest that parents may, in fact, make more restrictive decisions about the sharing of their child's data than about their own data (Burstein *et al.*, 2014; Patrinos *et al.*, 2022b).

In response to these considerations around participant informed consent, protective mechanisms can already be implemented, such as verification procedures for those applying for data access, and the requirement for applicants to submit a full research proposal and approved ethics application for analysis of secondary data to ensure that these correspond with the remit of research for which participants were originally consented (i.e., controlled access). Nevertheless, these protective mechanisms are predominantly coordinated by researchers, ideally in collaboration with stakeholder communities, and this emphasises the importance of trust building between research and participant communities to provide participants with confidence that their data will be managed and shared responsibly on their behalf, and for the foreseeable future. Of note, evidence from the current literature, also including the views of young people themselves (e.g., Pavarini *et al.*, 2022), suggests that public trust in research tends to differ according to context. For example, trust in local/national data sharing can be higher than in global data sharing; and trust in data sharing with industry is generally lower than with research and public bodies (Garrison *et al.*, 2015; Howe *et al.*, 2018b; Mello *et al.*, 2018; Pavarini *et al.*, 2022). Reduced trust in non-academic partners is important to better understand and address as international and industry expertise is often critical for the delivery of scientific advances in many areas of research.

Research quality and impact

Sample representativeness. Trust building between research and participant communities is not only needed from an ethical perspective, but also to ensure the highest research quality and impact (Kalkman *et al.*, 2019b; Rahimzadeh *et al.*, 2018). Indeed, if consent to data sharing is required to sign up to a research study, there is a concern that this may result in a skewed sample including only those who are willing for their data to be made accessible - possibly, therefore, further exacerbating the known bias in scientific research toward 'Western, Educated, Industrialized, Rich, and Democratic ('WEIRD') samples (see also Sanderson *et al.*, 2017). Similarly, where consent to data sharing is an optional component of study participation, there is a possibility that the 'public' dataset becomes less representative of the original sample from which it was drawn and thus provides fewer opportunities to replicate any findings from the original dataset. Equity and promoting fairness are key considerations here, since evidence suggests that trust in research differs between participant communities (Norstad *et al.*, 2022) and is associated with attitudes to data sharing (Matheny Antommaria *et al.*, 2018). For instance, Black, Asian and minority ethnic groups, educationally and/or socio-economically disadvantaged groups, looked after children, and those with intellectual disability, are among the paediatric participant communities most underrepresented in clinical research (National Institute for Health and Care Research, 2020).

Transparency over the purpose, importance, and nature of data sharing for research can promote willingness to opt in and consequently enhance the representativeness of publicly available datasets (Kalkman *et al.*, 2019b). To illustrate this point, we used data from our research consortium Autism Innovative Medicines Study - 2 - Trials (AIMS-2-TRIALS; https://www.aims-2-trials.eu/). We compared demographics and behavioural phenotypes of individuals who did vs. did not consent for their data to be externally shared (please see Table 2) as part of the AIMS Longitudinal European Autism Project (LEAP; for full details on the Methods underpinning LEAP and characteristics of the cohort, please see Charman *et al.*, 2017; Loth *et al.*, 2017). LEAP is an accelerated longitudinal developmental cohort study that includes autistic and neurotypical child, adolescent, and adult participants (average age 16 years; age range 6–30-years). Participants and parents/caregivers (where applicable) in the LEAP study were provided with details about the sharing of data with researchers outside of the consortium in the Participant Information Sheet and given the option to provide permission (or not) for the sharing of data in the Participant Consent Form. We note that there were no significant differences in consent preferences for external sharing when comparing minors (<16-years) and those 16-years and above ($p=0.34$, $\phi=0.03$; also see Extended data), therefore we report consent preferences for the whole sample here.

As shown in Table 1, consent to data sharing was provided by/on behalf of the vast majority (>80%) of LEAP participants. Moreover, there were no significant differences in questionnaire rated demographic and clinical characteristics of those who did vs. did not consent to external data sharing (please see Table 2).

Within our dataset, participants who agreed to sharing did not differ from participants who selected internal sharing only, emphasising that shared control of data curation between researcher and participant communities does not undermine data utility/quality. Analytical techniques, such as data imputation and simulations, can be employed to mitigate any remaining issues with missing data (Llera *et al.*, 2022). However, it is unclear whether a person who does not wish for their 'real' data to be shared would be comfortable with their missing data being imputed (e.g., see Randall *et al.*, 2021) and/or to be used to impute the missing data of others in the sample, and this remains a topic for further investigation.

Data degradation. Once a dataset is made publicly available, there remains a less commonly discussed tension between data sharing and other components of the open science movement, such as pre-registration. Pre-registration of hypotheses, protocols, and analysis plans before a study is conducted can increase the rigour of scientific inference and allow researchers to track the true 'false positive' rate across multiple uses of their data (Chambers & Tzavella, 2022). However, completely open datasets (i.e., those that do not require a project proposal submission/data access request, as opposed to controlled access) can be mined at will, making it challenging to determine whether an apparent pre-registration was lodged before any analysis commenced.

After an analysis has been conducted, it cannot be reversed, thus removing the possibility for other researchers to ask

Table 1. Proportion of AIMS LEAP participants who did vs. did not consent to data sharing beyond the consortium, by research location.

	Data Sharing 'Yes' (%)	Data Sharing 'No' (%)
UK	92% (N=330)	8% (N=30)
Netherlands	96% (N=277)	4% (N=11)
Germany	82% (N=55)	18% (N=12)
Italy	100% (N=42)	0% (N=0)

Note. Information in the table is based on consent preferences as recorded up to 20th February 2024. Participants were assessed at two timepoints, approximately 12–24-months apart. Figures reflect consent to external data sharing at both timepoints (sites in the Netherlands collected consent for both timepoints at timepoint 2), or timepoint 1 if the participant did not return for a timepoint 2 assessment (Data Sharing 'Yes'), or non-consent (including both active opt out and response not known) to data sharing at either or both timepoints (Data Sharing 'No'). UK sites include King's College London and the University of Cambridge. Sites in the Netherlands include Radboud University Medical Center Nijmegen and University Medical Center Utrecht. Sites in Germany include Central Institute of Mental Health Mannheim. Sites in Italy include Campus Bio-Medico University of Rome. Karolinska Institutet, Sweden was a LEAP timepoint 1 site but did not include a consent for data sharing item in the original Participant Consent Form and is therefore excluded.

Table 2. Comparison of participant demographics and characteristics by consent preference on selected key measures from the AIMS Longitudinal European Autism Project.

		Data Sharing 'Yes'	Data Sharing 'No'	Comparison		
				χ^2 (df)	φ	*p*
Demographics	***N* Group (Autistic:Non-autistic)**	417:287	33:20	0.19 (1)	0.02	0.66
	***N* Sex (Male: Female)**	494:210	36:17	0.12 (1)	0.01	0.73
				Z	***r***	***p***
	Age *M* (*SD*)	17.02 (5.83)	15.46 (6.63)	1.71	0.06	0.09
	FSIQ *M* (*SD*)	99.54 (19.82)	101.54 (20.83)	0.50	0.02	0.61
Autism features	**SRS-2 *M* (*SD*)**	64.93 (15.50)	62.76 (17.84)	0.78	0.03	0.43
	RBS-R *M* (*SD*)	12.45 (13.95)	11.77 (14.68)	0.56	0.03	0.56
	SSP M (SD)	149.84 (30.34)	150.85 (31.85)	0.16	0.01	0.87
Co-occurring features	**DSM-5 ADHD inattention *M* (*SD*)**	3.65 (3.31)	3.14 (2.82)	0.57	0.02	0.57
	DSM-5 ADHD hyperactivity *M* (*SD*)	2.20 (2.78)	1.57 (2.11)	0.66	0.03	0.51
	DAWBA Anxiety *M* (*SD*)	2.07 (1.29)	2.00 (1.25)	0.16	0.01	0.87
	DAWBA Depression *M* (*SD*)	0.70 (1.11)	1.14 (1.56)	1.01	0.04	0.31

Note. χ^2 (df)=Chi-square (degrees of freedom); φ=phi effect size for Chi-square test; p=p-value; Z=Z-value Mann-Whitney U test; r=effect size for Mann-Whitney U test. M=Mean; SD=Standard Deviation; FSIQ=Full scale IQ; SRS-2=Social Responsiveness Scale-2nd Edition; RBS-R=Repetitive Behaviour Scale-Revised; SSP=Short Sensory Profile; DSM-5 ADHD=Diagnostic and Statistical Manual 5th Edition Attention Deficit Hyperactivity Disorder Rating Scale; DAWBA=Development and Wellbeing Assessment. These measures were selected as they represent key features of interest in developmental autism research (the primary focus of the AIMS consortium) and are thus critical for sample representativeness.

targeted questions from a set of research data ('data exhaustion'). In addition, multiple uses of (and resulting publications from) the same dataset (Menon & Muraleedharan, 2016) (and/or the possibility that the same individual(s)' data are present in two or more combined datasets without the knowledge of the researcher) creates a dependency between hypothesis generation and hypothesis testing because the literature from which a prediction is derived is not independent from the dataset on which the prediction will be tested. These issues can result in data degradation (Hedley Thompson *et al.*, 2020), whereby the value of data declines as it is used, and thus pose important conceptual and practical considerations for researchers who create, curate, and analyse large-scale datasets.

Discussion: Some potential solutions

Taken together, the challenges discussed in the previous section align on the premise that data curation is essential to underpin optimal open science efforts in paediatric research; and that control of data curation should be shared between the participant and researcher communities.

Shared control of data curation

Dynamic consent is a key mechanism through which shared control of data curation between researcher and participant communities may be facilitated. Dynamic consent enables participants to actively adjust their consent preferences via a digital interface (including making advanced decisions, such as consent preferences over the use of their data after they have passed away), enhancing opportunities for researcher-participant interactions and engagement over time (Kaye *et al.*, 2015). This contrasts with more traditional one-time or broad consent models, whereby participants are asked to provide their consent preferences once at the start of a research project (Wiertz & Boldt, 2022) – often accompanied by very long (and thus less accessible) participant information and consent forms designed to 'cover all bases'.

In the context of paediatric research, an added advantage of dynamic consent over one-time consent is that it empowers young people to control their primary data sharing preferences as they become older, even if a parent/ caregiver initially provided consent on the young person's behalf. In a dynamic consent model, it is also possible for the young person to change their consent preferences without the need for the research team to individually recontact them, which can be time consuming, expensive (Budin-Ljøsne *et al.*, 2017), and sometimes impractical in longitudinal developmental research where participants can become uncontactable (e.g., contact details change over time) and/or research teams evolve (e.g., personnel change, teams disband).

The speed and efficiency of research progress and its translation to societal impact may also be improved (without compromising on participant choice, unlike other options, such as broad consent; Ferretti *et al.*, 2021b), as researchers are no longer required to individually recontact participants for re-consent as specific new research questions emerge. Limiting the need for multiple reconsenting procedures can further reduce burden on participants, depending on the granularity of the dynamic consent infrastructure used. For instance, models of dynamic consent that require participants to opt in or out of their data being used for new research projects every time they emerge may be perceived as burdensome (Steinsbekk *et al.*, 2013). However, dynamic consent could alternatively be implemented according to project type (e.g., 'all projects', 'anxiety research projects only') and/or project lead (e.g., 'all project leads (including non-academic)', 'academic project leads only'), location (e.g., 'global', 'Europe and US', 'Europe only'), or at the more general level of 'consent yes' and 'consent no' - with the critical component being that these preferences can be actively adjusted by the participant at any time.

Providing participants with true ownership over their primary research data in this way addresses the critical consideration of research purpose, emphasised earlier in this paper, and several studies now indicate this is also preferred by participant communities (Begum Ali *et al.*, 2023; Howe *et al.*, 2018a; McGuire *et al.*, 2011). As participants are given the opportunity to engage with ongoing research and dynamically opt in and out of the use of their data for addressing new research questions, researchers are provided with a 'barometer' of the value of diverse research topics to the relevant participant communities, therefore shaping research progress to increasingly align with the public interest. Furthermore, ongoing participant engagement and investment in the use of their research data could even improve the accuracy of that data – for instance, by providing participants with a clear and efficient communication mechanism through which they could update their research information if, for instance, they received a new diagnosis.

These prospective benefits of increased engagement with participants on the use of their research data contrast with potential concerns that participants frequently changing their consent preferences (and therefore altering datasets over time) will have a negative impact on research quality – a view that has not been supported by evidence to date (Teare *et al.*, 2021). For instance, participants involved in biobanking research do not necessarily envisage changing their consent preferences over time, though they would appreciate the opportunity to do so (Teare *et al.*, 2015) and research projects already utilising dynamic consent approaches (e.g., RUDY UK rare diseases study) do not report a high rate of participant consent withdrawals over time (Teare *et al.*, 2021). Nevertheless, with the use of dynamic consent, shared research datasets may alter over time (e.g., in the number of datapoints available for analysis, the data modalities available) and therefore research groups should provide summary information on the characteristics of the original sample vs. public dataset, and the release and version dates of the dataset should be clearly reported in published work to aid the interpretation of replication efforts.

To support individuals to make informed choices about the use of their data over time (particularly for minors, where data ownership is reliant on being notified about their data contribution), it is essential for research participants to have

access to educational resources around data sharing and for researchers to engage in effective communication with research participants, particularly regarding the purpose of data sharing. Indeed, poor communication around research purpose and use has been found to be a significant factor predicting participant non-consent (Norstad *et al.*, 2022), and this issue is becoming increasingly important as scientific approaches become more technical and thus more complex to convey to research participants in a way that supports them to provide fully informed consent on the use of their data. Strong communication includes providing accessible information on why the data are being collected, why and how the data will be shared, and transparency over the research groups' parameters for acceptable vs. unacceptable usage of the data (e.g., to align with the participant consent, and research groups' ethical principles). The latter is particularly important where data is fully open - in other words, where the research group does not maintain a level of oversight of data access, for instance through reviewing a project proposal submitted by those requesting access to the data. To further enhance transparency and support participant choice, those accessing the data should also be encouraged to also make their project proposal, ethical review, and results publicly available to provide a continuous record of research purpose and use.

Infrastructure and technical solutions

Despite the apparent promise of shared data curation, there are technical and logistical considerations that may explain why a high proportion of researchers do not (or cannot) ultimately share their data (Watson, 2022), and relatively few big data/ biobank studies currently utilise approaches like dynamic consent. One key issue is the technical infrastructure required to facilitate responsible and sustainable data sharing and the resources and costs implicated in this (Kaye *et al.*, 2015). For example, though technological advances can act as part of a solution for maintaining participant anonymity (e.g., the ability to 'de-face' MRI data; Bischoff-Grethe *et al.*, 2007)), human resources are still needed to prepare and organise data for sharing. This includes providing data dictionaries and metadata sufficient that other research groups can use and interpret the data accurately and coordinating communications with all those using a given dataset in the case of a consent withdrawal. In addition, the data repository used to host the dataset requires management (e.g., of participant consents, evolution of the data in live projects) and maintenance, with research groups oftentimes developing their own database infrastructure for their specific purposes. A disadvantage of this approach is that it reduces the interoperability of diverse datasets – one of the core principles of good scientific data management and stewardship (Wilkinson *et al.*, 2016).

A second key issue is navigating the logistical and legal considerations for sharing of data that is not, or cannot be, fully anonymised and/or is particularly sensitive. Legal frameworks, such as EU/UK GDPR are essential to protect individuals and their personal information, however, to realise ethical and sustainable data sharing frameworks, research communities need more support to implement these frameworks in a clear and timely manner (Vlahou *et al.*, 2021). Once again, we take the AIMS LEAP multisite study as an example, whereby there are seven primary data collection centres based in five countries across the UK and EU. Each site has their own local/national ethics review and interpretations/guidelines on data sharing – yet must agree on a single set of contractual terms for making research data available to an external party, and the external party(ies) (who may be based outside of the EU/ UK legislative frameworks e.g., in the US) must also agree to these terms. This relatively small-scale context already begins to illustrate some of the challenges for the progress and sustainability of data sharing, as research groups necessarily become more complex to bring together the international expertise required to address the most prominent research questions.

To support research communities to navigate these complexities, funders must provide specific resources and tools for research communities to deliver responsible and sustainable data resources. In particular, research communities need access to common and integrated resources, including secure, GDPR compliant database infrastructure for depositing data that can be made available beyond the lifespan of an individual research project, with harmonised, accessible templates for open science practices such as preregistration embedded within them. Such an infrastructure would be further enhanced by a centralised governance process (i.e., a rotating committee combining scientific and lived expertise from researcher and participant communities) to review and manage data access requests, according to the parameters for acceptable vs. unacceptable usage of the data provided by each individual research project/ group.

Access to a more universal framework within which responsible data sharing practices can operate optimally - also called for in other, related fields (e.g., please see Lautarescu *et al.*, under review; Rehm *et al.*, 2021) – will not only reduce duplication of efforts and total costs by saving each research group from developing a new/different process each time, but also increase global research equity (Evertsz *et al.*, 2023). For instance, though a fee structure can be implemented for data access requests to support ongoing curation once data is ready to be shared, initial funds are still required to first prepare and organise the data (Prameela, 2022) and these funds are often only available to larger, well-resourced research groups (see also Hajduk *et al.*, 2019; Tamuhla *et al.*, 2023 on equitable data sharing in global health research). These resources should be costed into a grant at the application stage and funders can support research communities by ensuring that advice and tools for data sharing are available to them from the outset of a project, so that adequate information can be provided to participants about where their data will be deposited before a study commences to facilitate informed consent.

Public engagement to pave the way forward

To conclude, data sharing has the potential to accelerate critical scientific advances, including in priority areas for global policy, such as paediatric mental health (World Health Organization (WHO), 2022). We are moving toward a significant shift in practice, where data sharing is expected, however the frameworks

within which this can and should operate optimally are not yet fully formed. In other words, data sharing is becoming 'required', before it is made possible and easy (Hajduk *et al.*, 2019; Nosek, 2019). This is perhaps particularly poignant in the context of child and adolescent psychiatry (and related fields) where there are somewhat unique circumstances around accessibility of research communications and consent management.

We make several recommendations for beginning to bridge this gap, with a focus on the development of a shared infrastructure, supported by funders (and public/governing bodies more broadly), within which research communities have equal access to the tools and resources needed for responsible and sustainable data sharing. This framework would provide participant communities with informed choice and ownership of their data and facilitate both research and participant communities to engage in ongoing dialogue and co-creation of solutions for managing data sharing that we can all be confident in (please also see primer on community engagement in research by McCabe *et al.*, 2024). To achieve progress and build trust, we need a public conversation about data sharing, given the potential of responsible data sharing practices to empower the general public to have oversight and ownership of research purpose and ensure accountability of research for the public benefit.

Ethics and consent

Full ethical approval for the original LEAP study was obtained at each study site (please also see Loth *et al.*, 2017) and participants (and parents, where applicable) provided full informed consent/assent. Of sites named in this article, specific details were as follows:

Site	Ethics committee	ID/ reference no.
In the UK		
KCL UCAM	London-Central and Queen Square Health Research Authority Research Ethics Committee	13/LO/1156
In the Netherlands		
RUNMC UMCU	Radboud universitair medisch centrum Instituut Waarborging Kwaliteit en Veiligheid Commissie Mensgebonden Onderzoek Regio Arnhem-Nijmegen (Radboud University Medical Centre Institute Ensuring Quality and Safety Committee on Research Involving Human Subjects Arnhem-Nijmegen)	2013/455
In Germany		
CIMH	UMM Universitätsmedizin Mannheim, Medizinische Ethik Kommission II (UMM University Medical Mannheim, Medical Ethics Commission II)	2014-540N-MA
In Italy		
UCBM	Universita Campus Bio-Medica De Roma Comitato Etico (University Campus Bio-Medical Ethics Committee De Roma)	18/14 PAR ComET CBM

Data and software availability

Underlying data

Underlying data cannot be shared for ethical reasons, as it is based on a comparison between those who did vs. did not consent to external data sharing in the AIMS LEAP project (meaning that a subset of the participants included here did not consent to their data being shared/made available). The data presented in this article have not been published elsewhere.

Extended data

Open Science Framework (OSF): Extended Data for "Data sharing in child and adolescent psychiatry research: Key challenges (and some potential solutions)" June 2024, https://osf.io/ujfnm/ (Oakley, 2024).

This project contains the following underlying data:

OSF_DataSharingPaper_ExtendedData_SupplementaryTable1_27062024.docx

Data is available under CC-By Attribution 4.0 International license.

Acknowledgements

We thank all LEAP participants and their families for their time and effort in participating in the study.

We also acknowledge the contributions of the AIMS LEAP Group: Jumana Ahmad, Sara Ambrosino, Bonnie Auyeung, Tobias Banaschewski, Simon Baron-Cohen, Nico Bast, Sarah Baumeister, Sven Bölte, Carsten Bours, Michael Brammer, Daniel Brandeis, Claudia Brogna, Yvette de Bruijn, Bhismadev Chakrabarti, Ineke Cornelissen, Daisy Crawley, Robert Dallyn, Cate Davidson, Flavio Dell' Acqua, Sarah Durston, Guillaume Dumas, Christine Ecker, Claire Ellis, Jessica Faulkner, Vincent Frouin, Pilar Garcés, David Goyard, Hannah Hayward, Joerg Hipp, Mark H. Johnson, Prantik Kundu, Meng-Chuan Lai, Claire Leblond, Xavier Liogier D'ardhuy, Michael Lombardo, David J. Lythgoe, René Mandl, Luke Mason, Andreas Meyer-Lindenberg, Carolin Moessnang, Laurence O'Dwyer, Marianne Oldehinkel, Bob Oranje, Gahan Pandina, Antonio M. Persico, Annika Rausch, Barbara Ruggeri, Amber Ruigrok, Jessica Sabet, Roberto Sacco, Antonia San José Cáceres, Emily Simonoff, Julian Tillmann, Roberto Toro, Heike Tost, Daniel von Rhein, Jack Waldman, Steve C.R. Williams, Caroline Wooldridge and Marcel P. Zwiers.

References

Bayley N: **Bayley Scales of Infant and Toddler Development, Third Edition (Bayley--III®).** *APA PsycTests.* 2005.
Publisher Full Text

Begum Ali J, Holman R, Goodwin A, *et al.*: **Parent attitudes towards data sharing in developmental science [version 1; peer review: 2 approved with reservations].** *Open Res Eur.* 2023; **3**: 182.
PubMed Abstract | Publisher Full Text | Free Full Text

Bischoff-Grethe A, Ozyurt IB, Busa E, *et al.*: **A technique for the deidentification of structural brain MR images.** *Hum Brain Mapp.* 2007; **28**(9): 892–903.
PubMed Abstract | Publisher Full Text | Free Full Text

Budin-Ljøsne I, Teare HJA, Kaye J, *et al.*: **Dynamic consent: a potential solution to some of the challenges of modern biomedical research.** *BMC Med Ethics.* 2017; **18**(1): 4.
PubMed Abstract | Publisher Full Text | Free Full Text

Burstein MD, Oliver Robinson J, Hilsenbeck SG, *et al.*: **Pediatric data sharing in genomic research: attitudes and preferences of parents.** *Pediatrics.* 2014; **133**(4): 690–697.
PubMed Abstract | Publisher Full Text | Free Full Text

Chambers CD, Tzavella L: **The past, present and future of Registered Reports.** *Nat Hum Behav.* 2022; **6**(1): 29–42.
PubMed Abstract | Publisher Full Text

Chan AW, Song F, Vickers A, *et al.*: **Increasing value and reducing waste: addressing inaccessible research.** *Lancet.* 2014; **383**(9913): 257–266.
PubMed Abstract | Publisher Full Text | Free Full Text

Charman T, Loth E, Tillmann J, *et al.*: **The EU-AIMS Longitudinal European Autism Project (LEAP): clinical characterisation.** *Mol Autism.* 2017; **8**(1): 27.
PubMed Abstract | Publisher Full Text | Free Full Text

Dalgleish T, Black M, Johnston D, *et al.*: **Transdiagnostic approaches to mental health problems: current status and future directions.** *J Consult Clin Psychol.* American Psychological Association, 2020; **88**(3): 179–195.
PubMed Abstract | Publisher Full Text | Free Full Text

Evertsz N, Bull S, Pratt B: **What constitutes equitable data sharing in global health research? A scoping review of the literature on low-income and middle-income country stakeholders' perspectives.** *BMJ Global Health.* 2023; **8**(3): e010157.
PubMed Abstract | Publisher Full Text | Free Full Text

Ferretti A, Ienca M, Sheehan M, *et al.*: **Ethics review of big data research: what should stay and what should be reformed?** *BMC Med Ethics.* 2021a; **22**(1): 51.
PubMed Abstract | Publisher Full Text | Free Full Text

Ferretti A, Ienca M, Sheehan M, *et al.*: **Ethics review of big data research: what should stay and what should be reformed?** *BMC Med Ethics.* 2021b; **22**(1): 51.
PubMed Abstract | Publisher Full Text | Free Full Text

Finn ES, Shen X, Scheinost D, *et al.*: **Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity.** *Nat Neurosci.* 2015; **18**(11): 1664–1671.
PubMed Abstract | Publisher Full Text | Free Full Text

Ford T, Mansfield KL, Markham S, *et al.*: **The challenges and opportunities of mental health data sharing in the UK.** *Lancet Digit Health.* 2021a; **3**(6): e333–e336.
PubMed Abstract | Publisher Full Text

Ford T, Mansfield KL, Markham S, *et al.*: **The challenges and opportunities of mental health data sharing in the UK.** *The Lancet Digital Health.* 2021b; **3**(6): e333–e336.
PubMed Abstract | Publisher Full Text

Garrison NA, Sathe NA, Matheny Antommaria AH, *et al.*: **A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States.** *Genet Med.* 2015; **18**(7): 663–671.
PubMed Abstract | Publisher Full Text | Free Full Text

Gooding P: **Mapping the rise of digital mental health technologies: emerging issues for law and society.** *Int J Law Psychiatry.* 2019; **67**: 101498.
PubMed Abstract | Publisher Full Text

Hahn T, Nierenberg AA, Whitfield-Gabrieli S: **Predictive analytics in mental health: applications, guidelines, challenges and perspectives.** *Mol Psychiatry.* 2017; **22**(1): 37–43.
PubMed Abstract | Publisher Full Text

Hajduk GK, Jamieson NE, Baker BL, *et al.*: **It is not enough that we require data to be shared; we have to make sharing easy, feasible and accessible too!** *BMJ Glob Health.* 2019; **4**(4): e001550.
PubMed Abstract | Publisher Full Text | Free Full Text

Hedley Thompson W, Wright J, Bissett PG, *et al.*: **Dataset decay and the problem of sequential analyses on open datasets.** *eLife.* 2020; **9**: e53498.
PubMed Abstract | Publisher Full Text | Free Full Text

Hoge SK, Appelbaum PS: **Ethics and neuropsychiatric genetics: a review of major issues.** *Int J Neuropsychopharmacol.* 2012; **15**(10): 1547–1557.
PubMed Abstract | Publisher Full Text | Free Full Text

Howe N, Giles E, Newbury-Birch D, *et al.*: **Systematic review of participants' attitudes towards data sharing: a thematic synthesis.** *J Health Serv Res Policy.* 2018a; **23**(2): 123–133.
PubMed Abstract | Publisher Full Text

Howe N, Giles E, Newbury-Birch D, *et al.*: **Systematic review of participants' attitudes towards data sharing: a thematic synthesis.** *J Health Serv Res Policy.* 2018b; **23**(2): 123–133.
PubMed Abstract | Publisher Full Text

Hunter D, Pierscionek BK: **Children, Gillick competency and consent for involvement in research.** *J Med Ethics.* 2007; **33**(11): 659–662.
PubMed Abstract | Publisher Full Text | Free Full Text

Insel TR: **Translating scientific opportunity into public health impact: a strategic plan for research on mental illness.** *Arch Gen Psychiatry.* 2009; **66**(2): 128–133.
PubMed Abstract | Publisher Full Text

Jao I, Kombe F, Mwalukore S, *et al.*: **Research stakeholders' views on benefits and challenges for public health research data sharing in Kenya: the importance of trust and social relations.** *PLoS One.* 2015; **10**(9): e0135545.
PubMed Abstract | Publisher Full Text | Free Full Text

Kalkman S, Mostert M, Gerlinger C, *et al.*: **Responsible data sharing in international health research: a systematic review of principles and norms.** *BMC Med Ethics.* 2019a; **20**(1): 21.
PubMed Abstract | Publisher Full Text | Free Full Text

Kalkman S, Mostert M, Gerlinger C, *et al.*: **Responsible data sharing in international health research: a systematic review of principles and norms.** *BMC Med Ethics.* 2019b; **20**(1): 21.
PubMed Abstract | Publisher Full Text | Free Full Text

Kaye J, Whitley EA, Lund D, *et al.*: **Dynamic consent: a patient interface for twenty-first century research networks.** *Eur J Hum Genet.* 2015; **23**(2): 141–146.
PubMed Abstract | Publisher Full Text | Free Full Text

Kirilova D, Karcher S: **Rethinking data sharing and human participant protection in social science research: applications from the qualitative realm.** *Data Sci J.* 2017; **16**: 43.
Publisher Full Text

Kozlov M: **NIH issues a seismic mandate: share data publicly.** Nature News, 2022.
Reference Source

Levenstein MC, Lyle JA: **Data: sharing is caring.** *Adv Methods Pract Psychol Sci.* 2018; **1**(1): 95–103.
Publisher Full Text

Llera A, Brammer M, Oakley B, *et al.*: **Evaluation of data imputation strategies in complex, deeply-phenotyped data sets: the case of the EU-AIMS Longitudinal European Autism Project.** *BMC Med Res Methodol.* 2022; **22**(1): 229.
PubMed Abstract | Publisher Full Text | Free Full Text

Loth E, Charman T, Mason L, *et al.*: **The EU-AIMS Longitudinal European Autism Project (LEAP): design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders.** *Mol Autism.* 2017; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Matheny Antommaria AH, Brothers KB, Myers JA, *et al.*: **Parents' attitudes toward consent and data sharing in biobanks: a multi-site experimental survey.** *AJOB Empir Bioeth.* 2018; **9**(3): 128–142.
PubMed Abstract | Publisher Full Text | Free Full Text

McCabe M, Lautarescu A, Heraty S: **Community engagement in research: a primer from UKRN.** February 17, 2024.
Publisher Full Text

McGuire AL, Oliver JM, Slashinski MJ, *et al.*: **To share or not to share: a randomized trial of consent for data sharing in genome research.** *Genet Med.* 2011; **13**(11): 948–955.
PubMed Abstract | Publisher Full Text | Free Full Text

Mello MM, Lieou V, Goodman SN: **Clinical trial participants' views of the risks and benefits of data sharing.** *N Engl J Med.* 2018; **378**(23): 2202–2211.
PubMed Abstract | Publisher Full Text | Free Full Text

Mello MM, Wolf LE: **The Havasupai Indian tribe case--lessons for research involving stored biologic samples.** *N Engl J Med.* 2010; **363**(3): 204–207.
PubMed Abstract | Publisher Full Text

Menon V, Muraleedharan A: **Salami slicing of data sets: what the young researcher needs to know.** *Indian J Psychol Med.* 2016; **38**(6): 577–578.
PubMed Abstract | Publisher Full Text | Free Full Text

Meyer MN: **Practical tips for ethical data sharing.** *Adv Methods Pract Psychol Sci.* 2018; **1**(1): 131–144.
Publisher Full Text

Miranda Montoya MC, Bravo Chamorro J, Leegstra LM, *et al.*: **A blank check or a global public good? A qualitative study of how ethics review committee members in Colombia weigh the risks and benefits of broad consent for data and sample sharing during a pandemic.** *PLOS Glob Public Health.* 2022;

2(6): e0000364.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Monteith S, Glenn T, Geddes J, *et al.*: **Big data are coming to psychiatry: a general introduction.** *Int J Bipolar Disord.* 2015; **3**(1): 21.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Mostert M, Koomen BM, van Delden JJM, *et al.*: **Privacy in Big Data psychiatric and behavioural research: a multiple-case study.** *Int J Law Psychiatry.* 2018a; **60**: 40–44.
PubMed Abstract | **Publisher Full Text**

Mostert M, Koomen BM, van Delden JJM, *et al.*: **Privacy in Big Data psychiatric and behavioural research: a multiple-case study.** *Int J Law Psychiatry.* 2018b; **60**: 40–44.
PubMed Abstract | **Publisher Full Text**

Munafo MR, Nosek BA, Bishop DVM, *et al.*: **A manifesto for reproducible science.** *Nat Hum Behav.* 2017; **1**(1): 0021.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

National Institute for Health and Care Research: **Improving inclusion of under-served groups in clinical research: guidance from INCLUDE project**. August 7, 2020.
Reference Source

Nature Biotechnology: **Data sharing in the age of deep learning.** *Nat Biotechnol.* 2023; **41**(4): 433.
PubMed Abstract | **Publisher Full Text**

Norstad M, Outram S, Brown JEH, *et al.*: **The difficulties of broad data sharing in genomic medicine: empirical evidence from diverse participants in prenatal and pediatric clinical genomics research.** *Genet Med.* 2022; **24**(2): 410–418.
PubMed Abstract | **Publisher Full Text**

Nosek B: **Strategy for culture change.** Center for Open Science, 2019.
Reference Source

Oakley B: **Extended Data for "Data sharing in child and adolescent psychiatry research: key challenges (and some potential solutions)" June 2024.** OSF. 2024.
https://osf.io/ujfnm/

Palk AC, Dalvie S, de Vries J, *et al.*: **Potential use of clinical polygenic risk scores in psychiatry - ethical implications and communicating high polygenic risk.** *Philos Ethics Humanit Med.* 2019; **14**(1): 4.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Patrinos D, Knoppers BM, Laplante DP, *et al.*: **Sharing and safeguarding pediatric data.** *Front Genet.* 2022a; **13**: 872586.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Patrinos D, Knoppers BM, Laplante DP, *et al.*: **Sharing and safeguarding pediatric data.** *Front Genet.* 2022b; **13**: 872586.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Pavarini G, Yosifova A, Wang K, *et al.*: **Data sharing in the age of predictive psychiatry: an adolescent perspective.** *Evid Based Ment Health.* 2022; **25**(2): 69–76.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Prameela SE: **Don't just mandate open data, fund it.** *Nature.* 2022; **602**(7898): 578.
PubMed Abstract | **Publisher Full Text**

Rahimzadeh V, Schickhardt C, Knoppers BM, *et al.*: **Key implications of data sharing in pediatric genomics.** *JAMA Pediatr.* 2018; **172**(5): 476–481.
PubMed Abstract | **Publisher Full Text**

Randall M, Stern A, Su Y: **Five ethical risks to consider before filling missing race and ethnicity data: workshop findings on the ethics of data imputation and related methods.** Urban Institute Racial Equity Analytics Lab. March, 2021.
Reference Source

Rehm HL, Page AJH, Smith L, *et al.*: **GA4GH: international policies and standards for data sharing across genomic research and healthcare.** *Cell Genom.* 2021; **1**(2): 100029.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ressler KJ, Williams LM: **Big data in psychiatry: multiomics, neuroimaging, computational modeling, and digital phenotyping.** *Neuropsychopharmacology.* 2021; **46**(1): 1–2.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ross MW, Iguchi MY, Panicker S: **Ethical aspects of data sharing and research participant protections.** *Am Psychol.* 2018; **73**(2): 138–145.
PubMed Abstract | **Publisher Full Text**

Sanderson SC, Brothers KB, Mercaldo ND, *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**(3): 414–427.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schneble CO, Elger BS, Shaw D: **The Cambridge analytica affair and internet-mediated research.** *EMBO Rep.* 2018; **19**(8): e46579.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Stein DJ, Shoptaw SJ, Vigo DV, *et al.*: **Psychiatric diagnosis and treatment in the 21st century: paradigm shifts versus incremental integration.** *World Psychiatry.* 2022; **21**(3): 393–414.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Steinsbekk KS, Kåre Myskja B, Solberg B: **Broad consent *versus* dynamic consent in biobank research: is passive participation an ethical problem?** *Eur J Hum Genet.* 2013; **21**(9): 897–902.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Stringaris A: **Editorial: *Trials* and tribulations in child psychology and psychiatry: what is needed for evidence-based practice.** *J Child Psychol Psychiatry.* 2014; **55**(11): 1185–1186.
PubMed Abstract | **Publisher Full Text**

Tamuhla T, Lulamba ET, Mutemaringa T, *et al.*: **Multiple modes of data sharing can facilitate secondary use of sensitive health data for research.** *BMJ Glob Health.* 2023; **8**(10): e013092.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Teare HJ, Morrison M, Whitley EA, *et al.*: **Towards 'Engagement 2.0': insights from a study of dynamic consent with biobank participants.** *Digit Health.* 2015; **1**: 2055207615605644.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Teare HJA, Prictor M, Kaye J: **Reflections on dynamic consent in biomedical research: the story so far.** *Eur J Hum Genet.* 2021; **29**(4): 649–656.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Tenopir C, Allard S, Douglass K, *et al.*: **Data sharing by scientists: practices and perceptions.** *PLoS One.* 2011; **6**(6): e21101.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Towse J, Rumsey S, Owen N, *et al.*: **Data sharing: a primer from UKRN**. Open Science Framework, 2019.
Reference Source

Vlahou A, Hallinan D, Apweiler R, *et al.*: **Data sharing under the General Data Protection Regulation: time to harmonize law and research ethics?** *Hypertension.* 2021; **77**(4): 1029–1035.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Walport M, Brest P: **Sharing research data to improve public health.** *Lancet.* 2011; **377**(9765): 537–539.
PubMed Abstract | **Publisher Full Text**

Watson C: **Many researchers say they'll share data — but don't.** *Nature.* 2022; **606**(7916): 853.
PubMed Abstract | **Publisher Full Text**

Watson E, Fletcher-Watson S, Kirkham EJ: **Views on sharing mental health data for research purposes: a qualitative study with people with mental illness.** *medRxiv.* 2022a; 2022.11.03.22281848.
Publisher Full Text

Watson E, Fletcher-Watson S, Kirkham EJ: **Views on sharing mental health data for research purposes: a qualitative study with people with mental illness.** *medRxiv.* 2022b; 2022.11.03.22281848.
Publisher Full Text

Wiertz S, Boldt J: **Evaluating models of consent in changing health research environments.** *Med Health Care Philos.* 2022; **25**(2): 269–280.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.*: **The FAIR guiding principles for scientific data management and stewardship.** *Sci Data.* 2016; **3**(1): 160018.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

World Health Organization (WHO): **World mental health report: transforming mental health for all**. 2022.
Reference Source

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 05 May 2025

https://doi.org/10.21956/openreseurope.19616.r53059

© **2025 Kirkham E.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Elizabeth Kirkham**
University of Edinburgh, Edinburgh, UK

This article discusses the value of sharing health data, addresses some of the ethical and logistical issues which occur when seeking to increase sharing of health data, and presents some potential solutions for meeting the challenges surrounding data sharing. The authors discuss their experiences of working with child and adolescent psychiatric data. This is a valuable focus, given that it covers data which are sometimes considered especially sensitive (mental health information, children's data).

Examining the demographics of people who did and did not consent to external sharing of data was a nice way to interrogate potential differences in people who do and do not wish to share their data. It was encouraging to see no apparent differences between the two samples, though firm conclusions are limited due to the small number of people who did not consent. In addition, as the authors note, the true differences between people who do and do not consent to data sharing are very difficult to probe, as these two populations likely differ substantially in their willingness to take part in research more generally.

The one suggestion for improvement I would make for this paper would be to include more discussion on the topic of low public trust in industry. This is addressed but not in substantial depth. It would be useful to consider how the solutions presented in the paper could be used to protect participants against potential risks from private-sector involvement in research, and/or whether additional solutions are required for this particular challenge.

I note that the recently released Atlas of Longitudinal Datasets, funded by the Wellcome Trust, may be able to contribute in some way to the solutions discussed in this paper. https://atlaslongitudinaldatasets.ac.uk/

This article makes a number of insightful points which are important for the field to consider. The statement "data sharing is becoming 'required', before it is made possible and easy" is highly prescient and important to acknowledge.

Is the background of the case's history and progression described in sufficient detail?
Yes

Is the work clearly and accurately presented and does it cite the current literature?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Not applicable

Are the conclusions drawn adequately supported by the results?
Yes

Is the case presented with sufficient detail to be useful for teaching or other practitioners?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Co-production in mental health research, mental health data sharing, neurocognitive correlates of mental illness and early life stress.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.