Commercial or industrial use of mental health data for research: primer and best-practice guidelines from the DATAMIND patient/public Lived Experience Advisory Group

Author	Affiliation(s) §	ORCID	Contributions
Jones, Linda A. *	Cam	0000-0001-9347-5715	Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Writing – review & editing
Barber, Corrine	LEAG	–	Investigation, Writing – review & editing
Clements-Brod, Naomi	Cam	0009-0009-9151-929X	Project administration, Writing – review & editing
Davies, Jan	LEAG	–	Investigation, Writing – review & editing
DelPozo-Banos, Marcos	Swa	0000-0003-1502-389X	Writing – review & editing
Hughes, Andrea	LEAG	–	Investigation, Writing – review & editing
Iveson, Matthew H.	EdB	0000-0002-7242-0456	Investigation, Writing – review & editing
Jain, Diksha ‡	LEAG	–	Investigation, Writing – review & editing
John, Ann	Swa	0000-0002-5657-6995	Conceptualization, Funding acquisition, Writing – review & editing
Maguire, Aideen	QUB	0000-0002-3932-2109	Writing – review & editing
Matins de Barros, Clara	LEAG	–	Investigation, Writing – review & editing
McIntosh, Andrew M.	EdP	0000-0002-0198-4588	Conceptualization, Funding acquisition, Writing – review & editing
McTernan, Michael	LEAG	–	Investigation, Writing – review & editing
Speechley, Jan	LEAG	–	Investigation, Writing – review & editing
Stewart, Robert J.	IoPPN, SLaM	0000-0002-4435-6397	Conceptualization, Funding acquisition, Writing – review & editing
Taylor, Jane	LEAG, OU	0009-0007-9835-7663	Investigation, Writing – review & editing
Ting, Louise	LEAG	–	Investigation, Writing – review & editing
Cardinal, Rudolf N. †	Cam, CPFT, CUH	0000-0002-8751-5167	Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing

* Author order is alphabetical except for the first and last authors.

† Corresponding author.

‡ Pseudonym; identity available to editor.

§ Affiliations:

- Cam = Department of Psychiatry, University of Cambridge, Clifford Allbutt Building, Bay 13, Cambridge Biomedical Campus, Cambridge CB2 0AH, UK.
- CPFT = Cambridgeshire & Peterborough NHS Foundation Trust, Fulbourn Hospital, Cambridge CB21 5EF.
- CUH = Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge CB2 0QQ.
- EdB = Centre for Clinical Brain Sciences, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ.
- EdP = Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF.

- IoPPN = Institute of Psychiatry, Psychology & Neuroscience, King's College London, 16 De Crespigny Park, London, SE5 8AF.
- LEAG = DATAMIND Lived Experience Advisory Group.
- OU = The Open University, Walton Hall, Kents Hill, Milton Keynes MK7 6AA.
- QUB = Centre for Public Health, Queen's University Belfast, Institute of Clinical Science, Block A, Royal Victoria Hospital, Belfast BT12 6BA.
- SLaM = South London & Maudsley NHS Foundation Trust, Bethlem Royal Hospital, Monks Orchard Road, Beckenham, London BR3 3BX.
- Swa = Swansea University Medical School, Singleton Park, Swansea SA2 8PP.

Version of: **5 Nov 2025**.

Keywords:

- health informatics
- health policy
- information management
- mental health
- patient and public involvement

CONTENTS

1 ABSTRACT....4
2 PLAIN-LANGUAGE SUMMARY....5
3 INTRODUCTION....5
4 MATERIALS AND METHODS....7
4.1 Research advisory group formation....7
4.2 Ethics approval....8
4.3 Scope and definitions....8
4.4 Guideline development....9
5 RESULTS....10
5.1 Glossary of relevant concepts, background knowledge, and previous guidelines....10
5.2 LEAG perspectives noted during guideline development....10
5.3 Guidelines....12
6 DISCUSSION....12
7 DECLARATIONS....15
7.1 Availability of data and materials....15
7.2 Declaration of interests....15
7.3 Funding....15
7.4 Acknowledgements....15
8 ABBREVIATIONS....16
9 TABLES....17
10 REFERENCES....31
11 FIGURE LEGENDS....42

1 ABSTRACT

BACKGROUND: Routinely collected health data, such as that held by the United Kingdom (UK) National Health Service/Health and Social Care (collectively "NHS"), has important research uses, but its appropriate use requires public trust and transparency. Commercial/industrial access to routinely collected health data is especially controversial and sensitive for the public, and particular concerns may relate to mental health (MH) data. Existing best-practice MH data science guidelines do not cover commercial uses specifically, but emphasise the importance of patient/public co-development of data science. **OBJECTIVES:** To develop patient/public-led guidelines for the commercial/industrial use of MH data for research, and to capture relevant background information required by patient/public participants. The focus was on the UK and its constituent nations, but the principles may have wider applicability. **METHODS:** A patient/public lived experience advisory group (LEAG) was set up within DATAMIND, the Health Data Research UK data hub for MH informatics research development. Initial training and discussion yielded a requirement for definitions and explanations of concepts and processes relating to MH data research, developed iteratively. Subsequently, the LEAG developed guidelines via a qualitative and iterative quasi-Delphi approach. The agreed scope excluded data provided for research with informed consent, data processing arrangements such as companies hosting electronic health records or e-mail systems on the instruction of health services, or compliance with legal minimum requirements. The scope included the use of routinely collected MH data (e.g. NHS data) for research by commercial/industrial organisations without explicit consent, and aspects of MH data collection directly by industry with consent. **RESULTS:** Alongside the primer in MH data research concepts, the LEAG provide recommendations and best-practice guidelines relating to commercial/industrial research use of MH data, for organisations controlling MH data (such as NHS bodies) and for commercial applicants seeking to use MH data for research. Alongside principles of transparency, patient rights, patient/public involvement in research, stringent governance, and statistical disclosure control, the guidelines recommend a risk–benefit approach to assessing applications for data use, within limits that include avoiding the export of unconsented patient-level data outside NHS-controlled secure data environments, and not providing access to unconsented free-text MH data to commercial applicants. We also provide some recommendations for NHS executive and regulatory bodies, relating to public choice and transparency, clarity of guidance to research-active NHS organisations, and support for de-identification. **CONCLUSIONS:** Patient/public involvement and understanding is central to MH data research. The primer materials developed here constitute information requested by public advisers prior to considering best practice. The guidelines reflect the views of people with personal or family experience of mental ill health. We hope they are of practical use to the wider MH research community and serve to increase public transparency and trust.

2 PLAIN-LANGUAGE SUMMARY

The NHS collects health information that can help with research. This includes information about people's mental health. But people often worry about how this data is used, especially if companies (not just doctors or researchers) want to use it.

A group of patients, carers, and members of the public worked together with mental health researchers to make new guidelines. These guidelines help make sure mental health data is used safely and fairly, especially by companies doing research.

The group learned about how mental health data is collected and used. Then they wrote guidelines based on what they thought was right and fair. These guidelines say that:

- *People should be told clearly how their data is being used.*
- *Companies should not get private notes (text) unless people have agreed.*
- *Mental health data should stay safe in NHS systems.*
- *Patients and the public should help decide how data is used.*

They also gave advice to NHS leaders to help protect people's privacy and make things clearer.

These rules were written by people with real experience of mental health problems. The aim is to protect people's rights and build trust while still helping important research to happen.

3 INTRODUCTION

Data collected and created as part of routine health care has an important part to play in health research, alongside data collected specifically for research purposes. Many health systems have mechanisms to use routinely collected data for research, as permitted by relevant legislation and research governance structures. In the United Kingdom (UK), the National Health Services (NHS) in England, Scotland, and Wales and Health and Social Care (HSC) in Northern Ireland (collectively "NHS") strongly support the use of anonymised, routinely collected health data for research without explicit consent (1,2), and the legal basis for such work is clear (3–6). A great deal of research is conducted in this way [e.g. (7)], and the importance of this field is likely to grow. For example, the UK Government has emphasised health data in its science strategy (8) and has been reviewing data protection legislation to support the analysis of data (9).

However, appropriate use of routinely collected health data also requires public trust, or a "social licence" (10). The processes governing research must be transparent and reflect patients' views (11–13). "Data science" can be complex, emphasising the importance of publicly accessible information about health data and how it is used (14).

One particular area of controversy or sensitivity is around commercial use of routinely collected health data. Such uses, and proposals for use, feature regularly in the UK media [e.g. (15)], and the UK government is

seeking to develop a pricing model for NHS data (16). There is public wariness about commercial access to healthcare data. This has been linked in part to limited public awareness about how data is typically used and safeguarded (17), though that is not the only reason (6). There is also considerable variation in what is meant by "commercial use". In the present work, we exclude the many common situations where the NHS pays a company to store or handle data, such as by providing computers, e-mail systems, or electronic health records systems, but where the NHS remains solely responsible for how the data is used. We consider instead situations where companies have active interests in the data for research. Even here, the spectrum is broad (discussed later). Past use of NHS data for research beyond healthcare has been particularly controversial, such as the use of de-identified hospital episode data, together with "geodemographic" profiling based on postcode, for actuarial (insurance) research (18–20). Our previous work has demonstrated strong UK public support for research using their de-identified health data by the NHS, academic institutions, and health charities; ambivalence around such use by profit-making companies researching treatments; and strong net opposition to access by other kinds of profit-making companies (6). The UK's national institute for data science, Health Data Research UK (HDR UK), supports several research hubs covering different disease areas, including the mental health (MH) hub, DATAMIND (DATA Hub for Mental health INformatics research Development), funded by the UK Medical Research Council (MRC). HDR UK has consulted the public about commercial access to data (21), set out plans to develop a commercial model for data access (22,23), and provided generic advice for considering commercial requests (24,25). There are operational commercial services for some HDR UK hubs (26).

The question of commercial access to data also relates to national data "pooling": assembling and linking data from many parts of a health service to improve service evaluation and/or research. In the UK, national data pooling was successful in 2020 onwards to support the response to the COVID-19 pandemic, relying on emergency regulations for the handling of data [e.g. (27)]. However, before this, the England-wide data pooling initiative "care.data" foundered (10,28) and another, General Practice Data for Planning and Research (GPDPR), has been paused (29–31). Concerns and uncertainty about plans for commercial access to NHS data played a significant role in the stopping of these initiatives (10,28,29), and soon after "care.data" was abandoned, Understanding Patient Data was set up as an independent body funded by a medical research charity, UK research councils, and government (32). The 2022 Goldacre review (33) recommended a small number of NHS trusted research environments (TREs) for hosting sensitive data, advocated a frank conversation about commercial use of NHS data for innovation, and noted the strong privacy safeguards that well-designed TREs can bring. From 2023, a number of UK regional NHS trusted research environments (TREs) or secure data environments (SDEs) are being set up (34), with plans for a national "federated" data platform (FDP) in England (35,36). The Faculty of Clinical Informatics have called for transparency about the FDP (37) in particular, given its scope. Data and Analytics Research Environments UK (DARE UK) recommended, after public consultation, that it should not matter who is using sensitive data (e.g. people working for academia, industry, or government) so long as the research is in

the public interest—a judgement that should be made involving the public—and the end-to-end process is transparent, with strong governance processes and rigorous vetting and monitoring of researchers (21). However, this consultation did not focus on MH data and we consider some other limitations in the present study. A related concept connected to TREs is that of Data Trusts, which are organisations designed to manage data for the benefits of the data subjects (38). The potential for Data Trusts is being explored within healthcare (39): patients do not legally "own" data about them held by services such as the NHS (40), but do have rights relating to the data (3).

There are reasons to consider MH data specifically when contemplating the commercialisation of routinely collected data or health data more broadly. MH data is considered more sensitive than some other forms of health data, slightly reducing (on average) people's willingness for it to be shared (6,41), and consistent with patients' experience of stigma attached to MH illness (42). Preferences are strongly influenced by the kinds of organisations that might receive data, and by how detailed the data is (6). Personal experience of MH illness is, however, associated with greater willingness (on average) to share MH data for research (6), and mental health research has been underfunded historically compared to many areas of physical health research (43). Best practice for MH data science involves incorporating the views of people with relevant personal experience (44,45). Therefore, in this study, we develop the views of DATAMIND's patient/public advisors into guidelines applicable to the commercial use of MH data. In doing so, we set out background information concerning health data research, particularly in the context of UK health services, as a primer that the patient/public advisors considered necessary in order to form their recommendations.

4 MATERIALS AND METHODS

4.1 Research advisory group formation

Funding for DATAMIND was awarded by the UK Medical Research Council (MRC) in May 2021 and commenced in December 2021. National advertising for the DATAMIND supranational (Super) Research Advisory Group (SRAG) membership was launched in November 2021 with a closing date in December 2021; the group was renamed the DATAMIND Lived Experience Advisory Group (LEAG) in 2025. Participants were sought with experience of mental ill health either personally or through a family member/friend, an interest in MH research using health data, and a desire to use their experience to shape research. Prior experience was not required; recruitment proactively sought a range of experience levels and emphasised that all perspectives were welcome. Interviews were held in January 2022, and 13 members were appointed, with 12 proceeding to the guideline development work. No applications were received from Northern Ireland, but all other UK nations were represented. LEAG members spontaneously reported diverse backgrounds, levels of research experience, knowledge about and views on secondary uses of data, personal experience of MH conditions and services, and protected characteristics (46). LEAG members were

remunerated for their time following National Institute for Health and Care Research (NIHR) guidelines (47).

4.2 Ethics approval

Ethics approval was not applicable (48).

4.3 Scope and definitions

The LEAG's focus was on the use of MH data by industry, given the particular sensitivity of this aspect of health data (6). The LEAG noted potential additional MH considerations around consent and vulnerability, and that relevant principles around the use of health data may frequently apply more generally, beyond the MH context. Recognising that individuals may in general consent to any use of their data that they so choose (3,4), if properly informed (49) and possessing capacity to decide (50), the LEAG did not focus on uses of health service data used for research with explicit subject consent, or data collected with consent within ethically approved research studies. Instead, they elected to consider primarily the use, without consent, of routinely collected MH data for research (e.g. research using NHS MH data), but also to consider issues relating to MH data collected directly by industry with consent, such as the nature of that consent process.

We use the terms "industry" and "commercial" to refer to for-profit organisations, typically companies, and any collaborations involving such organisations.

We did not consider truly anonymous data; under UK data protection law, this is not personal data, and by definition it is impossible to re-identify any person from it (3,4).

We did not consider companies involved in processing MH data but only as "data processors". Under European Union (EU) and UK law (3,4), data processors do not make decisions about the uses of data, but perform tasks as they are instructed by "data controllers", who are the "owners" and custodians of the data. For example, a company providing computing facilities, e-mail services, or electronic health record (EHR) systems to the NHS is likely to be a data processor handling sensitive confidential (fully identifiable) patient data. While such a company would be responsible for keeping the data secure, it would not be permitted to use the data for its own purposes or research.

While many of the principles, risks, and benefits discussed here would apply to data access by other "third-party" organisations, such as universities, medical charities, and other non-profit organisations such as social enterprises, there is also evidence that such organisations are seen more favourably in this respect by the public (6), and such entities were not the focus of the present work.

We also excluded all considerations of compliance with data protection legislation, applicable information governance rules, or patient/public rights in law, as these are a minimum mandatory standard.

4.4 Guideline development

Development was via a qualitative iterative quasi-Delphi approach (51). The DATAMIND academic team provided the LEAG with background materials and an induction programme (March 2022) that included teaching and question/answer sessions on types of health data, examples of the use of health data for research, existing MH data science guidelines, examples of prior work on public opinion on health data sharing, examples of industrial/commercial healthcare research, and examples of real-world controversies in health data use. This contributed to the iterative development of a glossary of relevant terms (see **Supplementary Methods**). The LEAG contributed to decisions around methods from this point. LEAG members were then embedded in the regular DATAMIND leadership and management meetings, with additional patient/public involvement (PPI) in the Strategic Advisory Board overseeing DATAMIND.

In May 2022, guideline development commenced. Academic and industry representatives briefed on the background and their perspectives. LEAG members contributed their views and perspectives to a recorded discussion; the academic team circulated minutes to the LEAG. The LEAG joined the regular DATAMIND Industry Forum (January 2023) to gain and give further perspective on the work of industry and, separately, advised on creation of a data literacy course by DATAMIND and McPin for the public. Some LEAG members attended additional meetings, such as with DARE UK and Research Data Scotland, bringing points and perspectives into development of these guidelines. The LEAG coalesced their views into a 1,950-word short-form note document summarising their considerations for sharing MH data with industry (with anonymity of all specific contributions). The main areas were (1) background knowledge and definitions; (2) the societal benefit of research versus risk of harm and potential mistrust of industry; (3) considerations on consent; (4) particular risks of negative impact on MH patients; (5) processes and other measures to control access to data; (6) security/privacy and identifiability issues; (7) data accuracy; (8) the process of guideline development by the LEAG; (9) routes to output for the LEAG's work in this area. A knowledge exchange session was delivered by the academic team based on LEAG-identified learning needs (October 2023). The academic team expanded these notes and views into the present paper (being one of the routes to output designed by the LEAG) with draft recommendations based on the LEAG views to date, definitions and explanations of terms as requested by the LEAG, and background information (including regarding health data science and regulations). A LEAG member reviewed existing industry-related health data guidelines, including those in the "grey" literature (e.g. web-based sources) (see **Supplementary Methods**).

The LEAG iteratively edited and revised this paper including the recommendations (from July 2023 to April 2025). Discussions involving the LEAG and industry members were held at the DATAMIND Industry Forum (8 December 2023), focusing in particular on how commercial/industrial research projects/researchers are treated by NHS organisations holding data, and how approvals for UK commercial/industrial MH research are and should be managed outside the public-sector health system. Alongside LEAG development of guidelines for organisations controlling data and for commercial applicants, additional recommendations

for NHS executive and regulatory bodies were drafted by academic members of the team with experience in health data research. Separately, the DATAMIND Business Development and Sustainability workstream produced a general guide to the MH data landscape for industry (52). A set of LEAG workshops were held on 8–9 July 2024 as part of finalising the guidelines. The LEAG decided not to rate/score individual guidelines; all were agreed, and they decided not to seek to prioritise some over others. After this process, all authors reviewed and edited the final manuscript, with further revisions discussed and agreed by e-mail, prior to submission.

5 RESULTS

5.1 Glossary of relevant concepts, background knowledge, and previous guidelines

The glossary of terms that was created is provided in the **Supplementary Results**. Some of the definitions thus created also formed part of the basis for the evolving online DATAMIND Glossary (https://datamind.org.uk/glossary).

The following topics of background information were considered of particular relevance and are presented here:

- What kinds of research is health data (including MH data) used for (**Table 1**)?
- How does data typically flow from health services to research (**Error: Reference source not found**)?
- What sorts of research do companies do (**Table 2**)?

Reviews of some existing guidelines were provided by a LEAG member, noting both strengths and limitations of content and the approach to guideline development, such as the level of patient/public involvement and of industry involvement (**Supplementary Results**). A common theme of existing guidelines was that MH data, and unstructured or free-text data, was often not considered explicitly (see **Supplementary Results**).

5.2 LEAG perspectives noted during guideline development

The LEAG noted some existing and potential benefits of health data research in general, alongside the principle of altruism via data sharing and the benefits and importance of de-identification/anonymisation. These included the potential for healthcare benefits via research, the potential for cost-effective research, and the potential for widening participation in research through the use of routinely collected data. They noted some existing/potential benefits of commercial/industrial health data research, including the potential for treatment developments and financial benefits to the NHS (**Table 3**). They also noted a number of concerns about this area, including concerns about health data research in general (such as data accuracy, concerns

around privacy and data security, security concerns, the need for trust and transparency, and the potential for misuse); additional concerns relating to MH data (such as additional sensitivity, the added potential at times for vulnerability when asked for consent, and a particular need for trust and confidentiality in therapeutic relationships); additional concerns relating to free-text MH data; and additional concerns relating to commercial access to data (such as diversity in companies' motivation and public opinion, previous examples raising public concern, and potential for additional difficulties in overseeing and regulating use) (**Table 4**). This list is not exhaustive (see e.g. **Supplementary Results**).

Variation was also noted in the financial structures surrounding commercial/industrial health research. Some commercial companies are traditional profit-based entities seeking to develop healthcare, e.g. pharmaceutical or medical device manufacturers; some may be non-profit entities. Benefits to public sector healthcare organisations may also vary, e.g. from free access to software developed from research, through "fee for service" in delivering research, through to participation in commercialisation of new products or technology.

The LEAG considered issues surrounding the process for overseeing and granting access to MH data for research, expressing views in the domains of oversight and accountability; ways to maximise transparency; the need to consider risks and benefits of proposed research (noting that public benefit is necessary but not sufficient, benefits and risks must be considered together, and while most health data research will not be completely risk-free, the approaches taken must seek to minimise risks, allowing only appropriate and necessary risks to be taken in a carefully safeguarded way); and on issues surrounding consent and opt-outs when applicable (**Table 5**). The LEAG noted that existing processes already include significant controls and these normally include PPI, such as when research is conducted under NHS research ethics approvals (53), when UK Confidentiality Advisory Group (CAG) approvals are sought for the use of identifiable patient information for research without consent (54), and within "local" TRE approval processes [e.g. (55)].

Outside the domain of research involving collaboration between industry and health services, the LEAG noted that "health apps", such as software applications available from public app stores, represent a relatively unregulated area—compared, for example, to software in use within healthcare services for decision support. The domain of "health apps", and of other online therapy and support services, is an area where data may be collected subject to "small print" consent forms, with relatively little clarity about final uses of the data; the LEAG deprecated such processes in favour of more transparency around data collection and use (**Table 5**).

The LEAG noted that one core principle is that that the anticipated patient/public benefit of research is fundamental to assessing it against any risks posed, such as to privacy. Another core principle is of trustworthiness and transparency, noting that coherent rules and guidelines assist this. The LEAG took the view that, where public benefits are strong, research is conducted transparently, privacy is strongly safeguarded, and the public are appropriately consulted, that potential industrial/commercial users of MH data should be treated by data providers in broadly the same way as health service/academic users, with the

exceptions that (a) free-text MH data should not be provided to industrial/commercial applicants, and (b) unconsented patient-level data should not be exported outside NHS-controlled secure data environments.

5.3 Guidelines

Recommendations were then produced, revised, and agreed (see **Methods**). The resulting specific recommendations and best-practice guidelines were divided into those for host (data controller) organisations and potential commercial users of MH data.

The guidelines for host organisations (data controllers, and potential data access providers, e.g. NHS bodies) are shown in **Table 6**. In summary, they emphasise general principles of transparency, patient rights and the need to support these via trained staff, PPI, the use of formal governance processes, and the need for awareness of regulatory and legal requirements. Individual applications for data access should be specified in detail and have their risks and potential benefits assessed by an oversight committee including PPI. The guidelines advise that the following be entirely avoided for commercial applicants: (a) high-risk models of data sharing, such as exporting unconsented data outside NHS-controlled environments; (b) the provision of unconsented free-text MH data, even if de-identified. Standard principles must apply as usual, including reseacher training and oversight of researchers, data minimisation, and statistical disclosure control before publishable results leave secure environments. Full details and additional points are in **Table 6**.

The guidelines for commercial applicants (potential data users) mirror those for host organisations, and are shown in **Table 7**. In summary, they emphasise the need for thorough awareness of information governance principles and applicable regulations (including an understanding of data collected with versus without consent for specific research purposes), involvement of patients and the public in commercial research, and training in statistical disclosure control. Fully specified protocols should be prepared for a project application to a hosting (data-controlling) organisation, and all requirements from that host organisation complied with, with transparency to the public throughout. Full details are in **Table 7**.

We also make recommendations for NHS executive and regulatory bodies, shown in **Table 8**. In summary, these cover the provision of central systems to simplify choice and consent from the public; clearer and consistent advice to data-controlling (e.g. NHS) organisations around aspects of research governance; support for efforts to de-identify data in situations where it is sometimes not de-identified at present; and continuing to improve transparency through citizens viewing their own health data, as well as through public visibility of research conducted with routinely collected data. For full details, see **Table 8**.

6 DISCUSSION

To our knowledge, there is no previous set of best-practice recommendations to guide UK public healthcare providers approached by companies about potential access specifically to MH data for research, nor for

companies seeking to make use of such data. In the present study, a research advisory group of patient/public members developed such guidance iteratively (**Table 6**, **Table 7**), and we make additional suggestions for executive/regulatory bodies (**Table 8**). We hope these are useful in practice, for these bodies and also for patients and the public, to increase awareness and expectations of good practice.

The guidance includes much that is uncontroversial: it is well established that there should be an emphasis on transparency, data safety and security, statistical disclosure control, and robust oversight (45). The guidance for host (data controller) organisations adds specificity, with organisation-wide guidance such as on staff training; ensuring a strong oversight regime is preferred (e.g. "research" over "service evaluation" approvals when in doubt); providing opt-outs; and avoiding exclusivity with industrial/commercial applicants (**Table 6**). For the consideration of individual proposed projects, the LEAG excluded some specific models of data sharing, namely the transmission of patient-level unconsented data outside NHS-controlled TREs and the provision of free-text data to commercial applicants, even when de-identified (**Table 6**). They set other models on a continuum of risk, advising strongly that lower-risk models should be preferred, and that higher-risk data access approaches can only be justified if there is a compelling public benefit argument and strong patient/public support for a specific project, which was envisaged to be exceptional. Linkage projects were noted to bring additional risks (**Table 6**).

An international study of young people's preferences around mental health data use has also supported the "secure server" or TRE/SDE framework (56), as did the Goldacre and Sudlow reviews (33,57). The Sudlow review notes the metaphor of SDEs as "reading libraries" rather than "lending libraries", with person-level data remaining in the secure environment (57). The recommendation to assess potential healthcare benefits versus risks, with detail on some relevant considerations (**Table 6**), reflects the public ambivalence about commercial research and the strong preference, within the sphere of commercial studies, for research with direct healthcare benefits (such as research into treatments). Net public preference is against data sharing to companies for other kinds of research (6). The emphasis on "public benefit" accords with another study noting that the public's concept of "public good" emphasises research that addresses real-world needs, including those affecting only small numbers of people; that addresses social inequity and inequality; and that minimises harm, e.g. by not perpetuating harmful stereotypes (13). Patient-led research is one way to focus research on these goals (58). The recommendations around transparency accord with further recent guidelines (59).

The additional concern around de-identified free-text data reflects, in part, the difficulties of "perfect" de-identification, which arises from the high level of detail, as well as sensitivity. Free text can have important research uses, both directly via clinical expert review [e.g. (60)] and by conversion to structured data via natural language processing (NLP) [e.g. (61)] before researchers are given access. The latter is one method to help safeguard privacy. A "safe setting" (TRE) is another important safeguard around managing free-text data itself (61), and the guidance reflects that (**Table 6**).

Guidance has perhaps been lacking particularly for commercial applicants seeking to use MH data for research, either via collaboration with healthcare provider organisations such as the NHS or for data collected directly with consent. The LEAG set out concrete recommendations for such organisations (**Table 7**), but also noted that the UK's health research ethics framework might usefully offer ethical review for some studies that may technically be exempt. An example might be if companies collect and study health data, with consent, direct from users (e.g. users of healthcare "apps"), with no reference to the NHS (**Table 8**).

The recommendations for executive/regulatory bodies (**Table 8**) also emphasise public information and citizen choice about data sharing and research participation, as well as the provision of clear and consistent guidance to NHS bodies seeking to conduct health data research. An emphasis on public communication is important for a number of reasons: transparency is one, and ensuring trustworthiness another, but potentially also willingness to contribute—the extent to which people agree that their data is valuable to society strongly predicts their support for commercial use of their de-identified health data for health research (19). Transparent communication and public awareness is also essential for methods involving opt-outs or where consent is taken as implied. The recommendation for a central mechanism to control clinical/research sharing of one's routinely collected data is consistent with the public's wish that the NHS should be in charge of, and administer the use of, patient data (62,63), and is an explicitly expressed public preference (6).

The strengths of this study include the development of practical guidelines in a somewhat controversial area of MH data science for which they were previously lacking; their development by members of the public with personal or other close experience of mental ill health forming a lived-experience research advisory group; and the inclusion of background information required by that group to consider the questions at hand and thus likely of use in broader public information work. The work also has limitations. While large-scale surveys can measure the ways in which their samples are biased relative to the entire population, and seek to correct for that to some extent [e.g. (6)], the present work involves a smaller number of people, whose views may not be fully representative of those of the UK populace or the subset affected personally by the issues considered. The non-quantitative methods represent in some ways a less rigorous approach to defining consensus than formal Delphi methods (45,51), though they enabled the capture of a broad-ranging set of background information requirements and individual specimen views (developed via rigorous methods) alongside the consensus recommendations, and without predefinition of statements to consider. The guidelines may require revision as use-cases and technologies develop, as for any guidelines.

Future commercial/industrial use of MH data has potential to bring benefits to patients but these potential benefits must be set against potential risks (**Table 3, Table 4**) (21,33). PPI is important when designing and considering individual research projects, but PPI can also contribute to framing an approach to this aspect of research in the round. We hope that the present guidelines help to do so.

7 DECLARATIONS

7.1 Availability of data and materials

Not applicable.

7.2 Declaration of interests

- AJ is a user of the SAIL Databank, a Wales-wide de-identified health research database (64), and principal investigator of the Adolescent Mental Health Data Platform (ADP) (https://adolescentmentalhealth.uk/).
- RJS is CI for the SLaM BRC Case Register ("CRIS"), a de-identified health research database (55,65).
- RNC consults for Campden Instruments Ltd in the area of research software (unrelated to the present work); receives royalties from Cambridge University Press, Cambridge Enterprise, and Routledge (unrelated to the present work); is the author of free de-identification/linkage software (66,67); and is CI for the CPFT Research Database, a de-identified health research database (https://www.cpft.nhs.uk/research-database).
- All other authors declare that they have no relevant interests.

7.3 Funding

Funded by the MRC (MR/W014386/1, MR/Z504816/1) via DATAMIND, the HDR UK mental health data hub. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission. This research was supported in part by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014, NIHR203312) and the NIHR Applied Research Collaboration East of England; the views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

7.4 Acknowledgements

We thank Mark Avery, Hannah Clarke, Sue Fletcher-Watson, Mike Garwood, Laura Hannah, and Kay Taylor for helpful comments, and Carrol, P.S., and two other LEAG members for their contributions. In memory of the late Dermot O'Reilly, of Queen's University Belfast, who also contributed to this work.

8 ABBREVIATIONS

ADP	Adolescent Mental Health Data Platform
ADR UK	Administrative Data Research UK (funded by the ESRC)
API	application programming interface
BRC	(NIHR) Biomedical Research Centre
CAG	UK Confidentiality Advisory Group (68). Created in 2013 to replace the National Information Governance Board (NIGB).
CBT	cognitive–behavioural therapy
CDA	confidential disclosure agreement (alternative name for an NDA)
CI	chief investigator
COVID-19	coronavirus disease 2019
CPFT	Cambridgeshire & Peterborough NHS Foundation Trust
CRIS	Clinical Records Interactive Search (software, or the resulting database)
DAC	Data Access Committee
DARE UK	Data and Analytics Research Environments UK (funded by UKRI, HDR UK, ADR UK)
DATAMIND	DATA hub for Mental health INformatics research Development
EHR	electronic health record (also: EPR)
EPR	electronic patient record (also: EHR)
ESRC	Economic and Social Research Council, part of UKRI
EU	European Union
FDP	NHS Federated Data Platform (England)
GP	general practitioner / general practice
GPDPR	General Practice Data for Planning and Research (England)
HDR UK	Health Data Research UK
HRA	UK NHS Health Research Authority
HSC	Health and Social Care (Northern Ireland)
ICO	UK Information Commissioner's Office (69)
LEAG	DATAMIND Lived Experience Advisory Group (formerly Super Research Advisory Group, SRAG)
MH	mental health
ML	machine learning
MRC	Medical Research Council, part of UKRI
NDA	non-disclosure agreement; see also CDA
NHS	National Health Service (England, Scotland, Wales) or HSC (Northern Ireland)
NIHR	UK National Institute for Health and Care Research
NLP	natural language processing
ONS	UK Office for National Statistics
PPI(E)	patient and public involvement (and engagement)
REC	Research Ethics Committee
RECOVERY	Randomised Evaluation of COVID-19 Therapy trial (70)
SAIL Databank	Secure Anonymised Information Linkage Databank (Wales)
SDC	statistical disclosure control
SDE	secure data environment (a broader term for a TRE)
SLaM	South London & Maudsley NHS Foundation Trust
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TRE	trusted research environment (see also SDE)
UK	United Kingdom
UKRI	UK Research and Innovation, a public body uniting the UK's seven research councils (which provide public funding for research)
US	United States of America

9 TABLES

Table 1: What kinds of research is routinely collected health data (including mental health data) used for? The table shows a few examples, but many studies use these techniques (71). For abbreviations, see **Abbreviations**.

Type	*Area*	*Example*
Epidemiological (observational) studies		
	COVID-19 risks and consequences	During the pandemic, the Secretary of State instructed NHS organisations to process identifiable health data for COVID-19-related public health purposes (27), due to exceptional circumstances and in a time-limited way. In one example, a company that provides electronic health record (EHR) software to many GP surgeries linked the primary care data it curates with NHS England data on COVID-19 infection and death, and a university team developed software to allow this data to be interrogated without researchers accessing the data directly (**Error: Reference source not found**). One of their first publications examined the risk factors for death from COVID-19, across 17.3 million people (72). This information helped decide the order of the UK's COVID-19 vaccination programme (73), a programme credited with preventing up to 128,000 deaths and 262,000 hospitalisations by September 2021 (74). Studies with different databases have examined mental health outcomes after COVID-19 (75).
	Blood pressure risks	Large-scale studies have accurately quantified the relationship between blood pressure and the risk of diseases such as stroke and heart attack (76).
	Green space and depression	EHR data has been linked to maps to show that having more green space close to home is associated with a reduced prevalence of depression and anxiety disorders (77).
	Cardiovascular disease in psychiatric disorders	Routinely collected health data has been used to measure the risk of cardiovascular disease in people with disorders such as schizophrenia and bipolar affective disorder (78), and risks relating to medications (79). Similar studies have shown that depression increases the risk of heart attacks (80,81).
	Lithium and dementia	Although bipolar affective disorder is associated with a slightly higher risk of developing dementia, lithium (a medication used for bipolar disorder and depression) has been associated with a reduced risk of dementia (82,83). Observational studies like these have been followed by randomised controlled trials (**Supplementary Table 1**) to see if lithium can slow the development of Alzheimer's disease; some have shown promise (84).
	Shingles vaccine and dementia	Recent studies with large-scale EHR data have shown that vaccination of older adults against shingles (herpes zoster, or varicella zoster virus, the virus that also causes chickenpox) reduces the risk of dementia (85).
Finding and recruiting participants for research		
	By clinicians	When asked to recruit patients for studies, clinicians may use EHRs to find potentially suitable participants —if they have an EHR, are permitted, and it's technically feasible (86). This includes mental health studies (87).
	By researchers	Some systems permit approved researchers, conducting ethically approved studies, to search de-identified databases (**Error: Reference source not found**) for potentially suitable patients. If the patient has previously consented (66,88), or if an electronic system can ask their clinician to approach them and they then consent (66), they can be identified to, and put in touch with, the researchers to consider participation ("consent for contact"). (Actual participation in a study requires further informed consent.)
Using routinely collected health data alongside newly collected research data		
	COVID-19 treatments	The RECOVERY trial recruited patients across the NHS, randomised them to potential treatments for COVID-19, and obtained routine health care data from their NHS records and national reporting systems. Amongst its discoveries, it found that dexamethasone reduced mortality amongst people with severe COVID-19 (70). Dexamethasone is widely available and very cheap, so this led to immediate changes in health care worldwide.
	Health anxiety	A study randomised people with health anxiety to a form of cognitive–behavioural therapy (CBT) or standard care, and measured symptoms (collecting new research data) and looked at health service use (via routinely collected data), finding improvements in symptoms (89).

Table 2: Examples of types of industrial or commercial research, meaning research conducted by or with companies or paid for by companies, relating to health or healthcare. Some examples involve companies acting independently and some involve companies collaborating with, or funding, universities or the NHS. For abbreviations, see **Abbreviations**.

Topic	*Comments*
Market research and internal research	Much research by companies is nothing to do with health care or the NHS. Companies might do simple "research" among their staff (e.g. seeing if they like a new mobile phone app), or market research (e.g. asking the public how much they would pay to use an app), without any requirement for ethical or regulatory approvals. There are industry standards for market research, however, including informed consent (90,91).
"Basic science" research	Similarly, they might do sophisticated biological or engineering research (e.g. on molecules or cells, in animals, or to build devices), some of which might need regulatory approvals, but which does not involve patients or the NHS.
Direct relationship with consumer	Companies may obtain data from people directly. For example, a company might invent a mobile phone app that allows users to record their mood. They might ask their users for consent to do research using their data, e.g. to see whether phone usage activity predicts sleep or mood. (By "consumer", we mean a user of a commercial product, including products provided "for free" by commercial companies.) Sometimes, one company might pay another to collect data.
Pharmaceutical and physical medical device research: clinical trials	When a pharmaceutical company invents a new medicine (e.g. a drug for cancer) or device (e.g. a heart pacemaker), it normally collaborates with healthcare (e.g. NHS) organisations, to test the invention in consenting volunteers. Typically, patients who wish to participate must consent to the company being able to see some of their health data for this, or the trial couldn't be done safely (because the company has to help oversee safety, including noticing if an unexpected side effect is happening across multiple sites).
Feasibility questions	Setting up a medicine trial is a lot of work, so pharmaceutical companies often ask multiple NHS organisations how feasible it would be to work with them, before committing. Part of that involves asking how many patients they see with the relevant condition. NHS organisations would typically provide a very basic answer to that question that cannot identify anyone (e.g. "50 per year").
Recruitment to clinical trials	If a commercial clinical trial starts, it needs to recruit participants. Sometimes the criteria are complex, e.g. "people aged 18–45 with depression plus a certain genetic marker who have tried at least three antidepressants and do not have epilepsy". Therefore, NHS organisations might use EHR data to "shortlist" patients to approach and ask about participation.
Computational health research: artificial intelligence and machine learning	A technology company developing software to "read" chest X-rays or to predict the risk of mental health crisis might want to train (teach) the software or test (validate) it using NHS data.
Epidemiological research	A company might want to know if a particular condition or problem is common ("how many people with bipolar affective disorder develop diabetes?"), or look for an unusual side effect ("how many people taking this antidepressant get serotonin syndrome and what are the risk factors?"). They might pay a university or NHS organisation to find out and publish the results. An EHR provider company might support NHS/academic research by enabling research using data that it looks after on behalf of NHS organisations (see COVID-19 example in **Table 1**).

Table 3: Potential strengths and societal benefits noted by the LEAG of commercial/industrial health data research. Some points are not specific to commercial/industrial use and relate also to principles of data use in general. For abbreviations, see **Abbreviations**.

Topic	*Comments*
Altruism and helping others	"If it could help someone then I would share."
Industry and industrial/NHS collaborations can create healthcare benefits	The LEAG noted the long history of companies creating health benefits, such as via drug discovery, the development or improvement of medical imaging technology, medical device development. These types of innovation may require healthcare data, sometimes from consenting participants. Companies have also worked to create healthcare benefits through the use of routinely collected data. These include decision support systems in EHRs, such as sepsis alerts (92,93), though these are not foolproof (94,95). De-identified data has been used to train machine learning (ML) systems to interpret X-ray images, such as for automatic cancer detection (96). The LEAG noted potential for healthcare improvement in this domain, noting also evidence that even simple changes to data provision can improve MH care (97). Companies that provide healthcare themselves have used routinely collected data to predict treatment response from psychotherapy (98–101).
Effective de-identification permits greater use	The LEAG endorsed the principle that effective de-identification or anonymisation of data can allow public benefits to follow, noting examples such as de-identified intensive care data sets available to approved researchers worldwide (102,103).
Full anonymisation permits public distribution of results	The LEAG noted the legal principle that truly anonymous data is no longer personal and does not need to be restricted (3,4,104), though this label should not be applied without careful thought (**Table 4**).
Widening participation	The use of routinely collected health data for recruiting to clinical trials and similar research, subject to further consent (**Table 1**), may allow greater and more diverse and representative participation in research.
Cost effectiveness	Use or re-use of existing data may reduce the costs associated with new data collection.
Scale	Use of large-scale data may allow new patterns to be identified in data, including via automated methods.
Potential financial benefits to the NHS and thus patients	"Will industry pay for data?"

Table 4: Concerns noted by the LEAG in relation to the use of health data for research: generically, relating to mental health conditions, and relating to industrial/commercial use. For abbreviations, see **Abbreviations**.

Context	*Topic*	*Comments*
Generic		
	Data accuracy	"Inaccurate data is a problem." "People should be able to see their own health data and correct errors." "Data should be usable and complete."
	Scope of data	"Concerns over industry getting more data than they need."
	Trust	Trust in organisations handling personal data has many aspects. These include being transparent; being well-intentioned and working for public/patient benefit; compliance with the law; holding data securely; not permitting identifiable data to be used inappropriately.
	Privacy and re-identification risk	The LEAG noted that there have been instances where supposedly anonymous data were susceptible to re-identification via jigsaw attack, even if this is rare (105,106). The threshold for "obvious" re-identification is one question (would a randomly selected citizen consider it identifiable?); the threshold for a malicious and mathematically sophisticated attacker is another (104,107,108). Because these questions are complex, the management of permissions to access NHS data without patients' explicit consent should involve input from people with expertise in the area of re-identification risk, to ensure privacy.
	Security	While the ICO regularly prosecutes a small number of health service workers, this is typically for accessing health (or other) records of individuals known to them, with the perpetrators having trusted access to a health records system but no lawful basis for access to those individuals' records (109). On a much larger scale, hackers have launched sophisticated cyber attacks on companies supplying computer services to the NHS, to break critical systems and hold data to ransom (110). As there are criminal incentives to obtain health data, data security methods need to be robust and resilient against attack.
	Transparency	The LEAG strongly endorsed the principle that the use of health data should be transparent throughout: methods, studies, and results (but not raw data) should be made publicly available. This principle of transparency applies equally to any data sharing with industry.
	Delays to benefit	While some "big data" studies and studies assisted by routinely collected health data have had immediate impact on public policy or healthcare [e.g. (70,72)], many have scientific merit but do not lead rapidly to such changes. The LEAG noted the need not to be overoptimistic about the impact, and time to impact, of the average study.
	Future law	The UK government has emphasised the importance both of NHS data and of the life sciences commercial sector (8), and has planned to make changes to data protection law (111). The LEAG noted the possibility that future weakening of data protection legislation might affect health data research.
	Future technology	Systems and methods need to keep up with new developments in technology, such as in the field of artificial intelligence, and new uses of data.
	Irreversibility of data use	Once results have been published, it is impossible for anyone whose data has contributed to opt out or withdraw consent. This usually applies to anonymous data, in which case there is no risk to any person, however undesirable they may consider it. Rarely, data would be identifiable or potentially identifiable (e.g. a case study with a patient as co-author or containing a high degree of detail), in which case this irreversibility is something that would be stressed during consent (112).
	Long-lasting effects of misuse	The LEAG noted that incorrect research or methods of concern can have a long-lasting deterrent effect. For example, fraudulent research linking the measles/mumps/rubella vaccination and autism increased rates of vaccine refusal (113–115). Proposals interpreted as making NHS patient data available to private companies were followed by widespread opting out of NHS data sharing (116).
Mental health		
	Additional sensitivity	There is substantial research showing that mental health data is one category of data viewed by patients as being particularly sensitive. However, some quantitative studies have found that this concern is significantly less than other factors such as data destination, and people with personal experience of MH illness were significantly more willing to share their de-identified MH data for research than people without (6).
	Free-text data	Particular sensitivity was noted around the use of free-text MH data, which can be exceptionally detailed and sensitive.
	Vulnerability when consenting	"When unwell you are not always fully aware of what you are doing—you may give data you later regret. It's possible to get data from vulnerable people without them understanding the risks." The use of permissive data sharing policies by MH crisis charities has been cited as one example (117).
	Distress	"[This] can cause a lot of distress, which can worsen health."
	Trust in therapeutic relationships	"It takes a lot of courage to open up in therapy, and 'taking data' can prevent openness in therapy, which could be detrimental to clinical care. The relationship with one's therapist is important for effective care, and violation of trust (e.g. by taking data without consent) can impact this."
	Misuse	"Can it be used against me in the future?" The LEAG noted concerns around the potential for (a) identifiable data, given with consent, to have implications such as for insurance, and (b) de-identified data to be used for ethically debatable purposes, such as genetic studies being used as the basis for prenatal screening for conditions (e.g. in mental health).
	Perceived lack of benefit for MH services	While many research studies aim to deliver future clinical benefit, the actual state of MH care in the UK has deteriorated, as perceived by patients (118) and clinicians (119) (see also "delays to benefit" above).
Industrial/commercial		
	Privacy	"There is a fine balance between helping others (by sharing data) versus privacy. Controlled access is required."

Motivation	"How can we be sure companies aren't doing this for financial gain only, with no public benefit?"
Diverse public opinion	A minority, but a significant minority, do not want their health data shared at all, and sharing to industry is viewed much more negatively than to many other destinations (6). The LEAG noted that is therefore critical that, at the least, opt-outs are available and widely publicised.
Control over use	The LEAG noted that if data flows to third parties such as companies, it may be harder for data owners to verify actual uses against approved uses, so robust mechanisms are needed to ensure oversight.
International regulation	"What happens if data goes abroad?" "How can we be sure they have the same security rules?" "I would not want to share outside of the UK." These LEAG concerns related to patient-level data; it was accepted that truly anonymous data are suitable for worldwide publication and are routinely published without any restrictions on distribution. The 2023 UK–US Data Bridge was noted (120); the UK Government has indicated that this is intended to permit UK personal data to flow to the US while providing legal protections to UK citizens, and it may be used for health research (121). However, concern remained about the potential for weaker regulatory requirements in other jurisdictions, if potentially identifiable data were allowed outside its hosting (data-controlling) organisation.
Trust and previous uses raising public concern	"Tired of hearing 'you can trust me'." The LEAG noted a well-publicised research-related NHS breach of data protection law (122–124), though one not severe enough to warrant damages (125). They noted that central NHS initiatives around patient data transfer that have foundered partly for lack of clarity about the extent of potential future sharing and use, such as "care.data" (28) and GPDPR (29–31). Charities have shared de-identified text from MH crisis support services with academic and commercial organisations, provoking criticism despite privacy policies allowing such use (117). Cyberattacks involving NHS data continue (126).

Table 5: LEAG opinions and perspectives on the MH data access request process for research. For abbreviations, see **Abbreviations**.

Context	*Topic*	*Comments*
Oversight		
	Patient and public involvement	"Need to have PPI both for the data providers and in industry."
		"Always have a glossary, preferable designed or co-designed by patients and the public."
	Guidelines	Guidelines should be clear to all involved. All processes related to data sharing should be clearly documented in plain language.
	Accountability	"There should be clear accountability if something goes wrong."
Transparency		
	Personal access to medical records	"Do people know what is in their medical file?"
		"People should be able to see their own health data."
	Staff training	"NHS staff should understand how data is used."
	Public information and engagement around research	"There should be information about how data is used, and support to help people decide about the uses of their data."
		"Manage expectations": patients should expect that their data be used anonymously for research. This pledge is already part of the NHS Constitution for England (1,2) (**Supplementary Table 3**).
		"Public, widespread information campaigns should be used if there is an intention to collate NHS patient data [from multiple sources]."
		The results of research using patient data should be publicly accessible. [Prompt open access to the results of research is already required by the major UK health research funders (127–129).]
	Verification of outputs	Results and publications should be linked to their project approvals, via public registries, to enable readers to check adherence to approved applications.
Risk		
	Risk–benefit balance	"Access by industry isn't black and white, with some processes always being OK and some always being wrong. It's a risk/benefit trade-off. As data gets more sensitive, the anticipated public benefit of the research should be correspondingly greater [to justify access]."
	Default of patient-level data remaining in NHS TREs	A default implementation should be that patient-level data from NHS services should not leave NHS environments (or trusted third parties appointed by the NHS to manage its data securely in TREs), because such data may be vulnerable to re-identification via jigsaw attack.
	Additional sensitivity around data linkage	When data from multiple sources is linked, there is additional detail that might support jigsaw attack (**Supplementary Table 2**), and therefore the linked data is highly likely to be intrinsically more sensitive. In addition, data linkage processes should be designed so as not to provide extra information to the organisations linking data. [For example, when health and education data is linked for research, the education provider should not learn health information about its students, or vice versa (130).]
	If possible, isolate researchers from patient-level data	Prefer automatic statistical disclosure control (SDC), with researchers separated from data, to direct researcher access to data (**Error: Reference source not found**) (33). Recognise that automatic SDC remains imperfect.
	Minimise data scope	Require researchers to define people, variables, and time spans of interest. (If the data provider is creating a custom "curated" data set for a study, restrict to the data requested. If researchers are provided with access to a larger data set within a TRE as a matter of routine, audit that they are only using the data requested—but this overall approach was deprecated relative to data subset provision.)
	Free text	Free-text MH data represents a quantity and level of detail of sensitive data that should not be shared to commercial applicants.
	Consider different industry–NHS models	For any given level of detail of the data (e.g. level of risk of inadvertent re-identification), the LEAG perceived different ways of NHS–industry collaboration as falling on a further spectrum of risk. Models of collaboration are presented in **Table 6**.
Consent and opt-outs		
	Informed decision-making	"Informed decision-making requires people to understand what their data looks like and how it is used" (as above).
		The LEAG noted that the framing of questions can influence decisions around data sharing, perhaps to a small degree (6). It considered that it was desirable for NHS bodies or others to be supportive of research in general, but not to be coercive to patients in any way, and that the NHS must be clear that decisions about research must never affect the standard of care they receive.
	Opt-out process for routinely collected data	"Some people want to make a decision each time, but it isn't feasible to seek consent for every single study." Most of the LEAG, but not all, endorsed an opt-out process. The LEAG noted that opt-outs may not be required legally for de-identified/anonymised data, and may bias research against groups who are more likely to opt out (131), but considered that opt-outs are normally desirable unless there are strong public interest reasons otherwise (such as research specifically to benefit such groups).
	Level of detail around opt-outs	The biggest factor determining the public's willingness to share their health data for research is its destination, but detail (free-text versus structured data) is also important (6). Some people would like to have control over other specifics aspects of their data (such as mental versus physical health, sexual health or alcohol/substance use versus other kinds of data, or historical versus recent data) but a majority supported a specimen system that did not distinguish these subtypes (6). The LEAG took the view that opt-outs should be available and publicised, including for de-identified data unless there

	were to be a compelling public interest reason to override an opt-out, but that further fine-grained detail on the types or uses of data should not be mandatory.
Local decision-making by NHS bodies about routinely collected data, with patient/public involvement	The LEAG noted that the relationship between local and central NHS bodies is sometimes opaque to the public. It recognised the importance, for some research, of centralised data pooling and management (33). It strongly endorsed the principle of public clarity around both central and local data use, and supported the principle that local NHS organisations should be able to support locally relevant research, subject to necessary approvals, in addition to centrally organised research.
Informed consent ("opt in") for direct participation and where data is newly collected for research	Studies collecting new data from patients/public must obtain informed consent. Consent processes (e.g. patient information materials) should be reviewed prospectively by members of the public. NHS Research Ethics Committe (REC) approvals incorporate this review, but it should also be used by companies collecting MH data from the public. "Small print" consent forms should be avoided. In NHS studies, consent is normally taken face to face by a researcher with training in mental capacity assessment, which should reduce the risk of taking consent inappropriately from someone who is currently too unwell to consent. If companies collect bulk data online, however, this mechanism is absent; consequently, consent should be revisited regularly in that context (see below).
Prospective data use	Researchers collecting new data should be explicit about future data use, regardless of the scenario. Data given by patients and the public is valuable and patients are likely to expect it to be used for maximal research benefit. If envisaged, this should be made clear from the outset, reducing ambiguity or the need to attempt to re-contact and re-consent participants.
Revisiting consent	Companies collecting data over a prolonged period (e.g. months or more) should regularly re-seek consent. This is particularly important for automatic data collection (e.g. via mobile phone apps).
Withdrawal of consent	While data is identifiable or pseudonymised, there should be mechanisms for subjects to withdraw consent and have their data removed. (It is not realistic to expect data that has proceeded to an anonymous stage to be removed by repeating all research work to regenerate new anonymous results.)
Verification of data and access to data	UK participants should be made aware that they have a right to access identifiable data about them (3,4). There are some restrictions on access to medical records but not for data held in most other contexts, including when companies hold personal data. Citizens can use this right as a means of checking (for example) that data held by a company is correct, or has been deleted where applicable.

Table 6: Guidelines for organisations hosting (H) mental health data as data controllers, i.e. potential data access providers (such as NHS organisations), approached by industrial/commercial applicants relating to MH data requests for research. For scope, see **Methods.** For abbreviations, see **Abbreviations**.

Category	*Number*	*Summary*	*Detail*
General			
	H1	**Be transparent about research and data use**	Public engagement, transparency, and openness about both the benefits and risks of research are important to earn the trust of patients. Organisations should have a strategy for communicating their intended uses of health data, people's rights, and how to contact staff for more information. This includes being transparent about any provision of data access to commercial organisations. Criteria for access to data, and documentation of all processes for using data, should also be explicit, transparent, and available online. Communications should be in plain language and via multiple accessible routes.
	H2	**Train patient-facing staff**	Clinical staff interacting with patients may need to express clearly the uses of health data, either for direct care (132) or for research. This requires at least basic training for all such staff, but it is likely better for NHS organisations to have a highly trained subset of staff whose function is to support research in the NHS and who can provide in-depth support to patients with queries.
	H3	**Maintain, train, and embed a patient/public advisory group**	Organisation-specific end-to-end patient and public involvement (PPI) is essential, incorporating diverse views of people with relevant experience. Equality, diversity, and inclusion should be considered throughout. In addition to diversity of protected characteristics recognised in law (46), consider diversity of geography (e.g. rural/urban communities), socio-economic status, and views about data sharing, whilst recognising the impossibility of full representativenes in small groups. Do not exclude critical or sceptical voices. PPI groups need training in relevant aspects of data science and research oversight. The organisation's research data strategy, formal processes, and communications should be co-developed (co-designed) with PPI input.
	H4	**Emphasise patient rights**	People's decisions about research must not affect the standard of care they receive, and they should be made aware of this.
	H5	**Prefer "research" governance over "audit/service evaluation"**	The governance rules for research (typically regulated in the UK via ethics committees) are more stringent than for NHS audit/service evaluation (designed for NHS organisations to evaluate their own services and typically approved "in house"). Use of a "service evaluation" route for work beyond its proper scope has been associated with breaches of UK data protection law by the NHS in the past (122–124). Work involving external applicants is likely to be research, not service evaluation. If in doubt, use the research governance route. (See also **Table 8**.)
	H6	**Seek REC review for NHS research databases**	This may not always be mandatory at present (see **Table 8**) but it is strongly recommended.
	H7	**Offer opt-outs**	Advertise patient data rights and, where applicable, the NHS National Data Opt-Out (**Supplementary Table 4,** recognising its scope is restricted). In addition, provide and advertise local opt-outs from the use of de-identified data for research.
	H8	**Avoid commercial exclusivity**	No commercial organisation should have exclusive or privileged access to research opportunities or data (33).
	H9	**Review progress and procedures regularly**	Law, regulations, and best practice evolve rapidly in this field. So do technological advances and security threats. Organisations should have a designated person accountable for oversight of data security/access. There should be a structure to ensure that formal procedures and controls are reviewed regularly, and updated as necessary, with PPI, including learning from any failures, misuse, or breaches. Formal processes should include the response to data breaches and/or rule breaches, including investigation, remedies, and sanctions.
Specific to individual projects (data access requests)			
	H10	**Avoid secrecy**	Reject non-disclosure agreements (NDAs) surrounding access to healthcare data. Consenting research participants should never be asked to sign NDAs. Provider organisations should not sign NDAs relating to the provision of access to data without patient consent. NDAs may be acceptable about other aspects of a project (e.g. non-disclosure during scoping) (133) but view them with scepticism, as NDAs can conflict with being transparent.
	H11	**Require applications to be clear and detailed**	Applications to access data, and applications to amend previously approved projects, should be formally documented, made via a formal process, and assessed formally. Require applicants to provide a detailed protocol that includes full details of the purpose of the study and the data requested, such as via the STROBE format for epidemiological studies (134). Every application should have this, even if the work is conducted under overarching approvals for a research database. Applications should specify the lead researcher, all other researchers (i.e. everyone who would access data), and their organisation(s). Protocols should include a plain-language summary.
	H12	**Embed PPI**	All applications should be reviewed by members of the provider's patient/public community as part of the organisation's oversight and governance. Seek views from a broad and diverse range of public contributors. Additional PPI input into the proposal prior to the application is desirable, and should be described by the applicants. To enhance the relevance, impact, and inclusivity of research it is recommended that researchers aim to include public co-applicants in their projects, which requires funding, training, and support. Embed PPI into all aspects of governance and oversight of research studies.
	H13	**Assess applications using a risk–benefit approach,**	An oversight group (Data Access Committee; DAC), including patient/public representation, should consider the likely benefits and risks (in conjunction) of all proposed projects. The DAC should have oversight of the entire data lifecycle: collection, storage, access, and use (135). Use clear, documented assessment criteria. Consider applications according to the Five Safes model (136,137).

	within limits	**Benefits**—these will be specific to each project but consider in particular: ► Require a clear and material potential public benefit before permitting commercial studies, recognising their increased sensitivity (6). Public benefit must be the primary aim of the project; this does not preclude profit, but not all profit-making initiatives have potential for public benefit. Non-profit social enterprises (138) are likely to be viewed more favourably by the public, as are health charities (6). Among companies, the public have a clear preference for research with likely direct healthcare benefits, such as research into treatments, over other kinds of commercial research (6). Public benefit is necessary but not sufficient. **Risks**—consider in particular: ► *Re-identification and privacy.* • The use of directly identifiable data is obviously the most sensitive; using identifiable data for research without consent requires special national approvals, and must not be done without this. • With de-identified data, more data and greater detail (including linked data) increase the potential re-identification risk. Re-identification might be inadvertent or deliberate; without permission, the latter is criminal (3), but the risks should be minimised regardless. Input should be sought from people with appropriate expertise in assessing re-identification risk. • De-identified free-text data is an obvious example where greater detail increases risk, and mental health data is both detailed and highly sensitive. **Do not provide access to unconsented free text MH data, even if de-identified, to commercial applicants.** (Consider instead using NLP to produce structured data from that free text, if required.) • Preventing researchers from having direct access to the data, with automatic statistical disclosure control (SDC), is at the safer end of the spectrum (but recognise that automatic SDC may be imperfect and SDC processes require human verification). ► *Level of access.* On a spectrum of risk, specimen models of NHS–industry collaboration include: 1. In model 1 (lower-risk end), a company would have an "arm's length" role in which it commissions another body (e.g. an NHS Trust or university) to conduct the research, with results being publicly available, including to the commissioning company, but without the company having any direct access to data or control over queries. 2. In model 2, companies could be granted access to NHS-controlled systems that perform automated SDC (**Error: Reference source not found**), allowing them to perform queries for approved projects, but without direct access to data. 3. In model 3, named industry researchers could be granted access to de-identified data within NHS-controlled secure research facilities, in the same way that NHS/academic researchers often are. Within this model, access from monitored "safe rooms" [e.g. (139)] is preferred. 4. **Avoid:** In model 4 (high-risk end), de-identified NHS data would be exported to an industrial secure research environment, with NHS oversight. **This model should not be used for unconsented patient-level MH data;** see also **H18** below. **Lower-risk models should be preferred. Model 4, or anything less stringent, should not be used.** ► *Misinterpretation and inaccuracy.* Consider whether the project has sufficient academic and clinical expertise, "in house" or via collaboration, to ensure that interpretations drawn from the data are justified. ► *Potential for misuse, including inadvertent mishandling of data.* Check applicants' and their organisations' experience and reputation, including in handling sensitive data. Examine formal certifications [e.g. Accredited Researcher status (140,141)]. See also **H19** below. ► *Loss of trust or reputation.* Consider the implications of the project for trust in your organisation and its services. This list is not exhaustive. **Decisions**—Expected decisions by an oversight committee might be: approve, approve with modification, modify and resubmit, or reject.
H14	**Ensure researchers are trustworthy and trained**	Vet researchers and their organisations. Ensure all researchers have training in information governance and statistical disclosure control, and this training is kept up to date (see also **H13** and **Table 7**).
H15	**Minimise data provided**	Provide only the minimum data required to address the research topic. Require researchers to define people, variables, and time spans of interest (see **H11** above). Do not provide free-text data to commercial applicants (see **H13**).
H16	**Restrict and monitor data access and use**	Implement technical methods to restrict researchers' access to data. This might be by creating a custom data subset for a study, or equivalent technical mechanisms. Audit that they are only using the data requested, and according to the protocol. Consider mandating a high-security monitored environment such as the SafePod network (142).
H17	**Ensure statistical disclosure control**	Implement methods to check outputs for safety (i.e. that they are non-disclosive) prior to publication. This is particularly important when manual SDC is in use (**Error: Reference source not found**).
H18	**Keep data "in house" when at patient level**	Use a data handling model that keeps patient-level de-identified data within NHS-controlled secure environments (see also **H13**), unless the data is truly anonymous (poses no re-identification risk). ("NHS-controlled environments" may include commercial computing providers acting solely as data processors to provide the NHS with high-performance computing. See also above and **Table 5**.)
H19	**Set out contractual obligations with**	A contractual relationship should be put in place. This should address and include: (a) data flow, including prohibitions on redistribution or further data sharing, attempted re-identification, unauthorised linkage with other data, other use beyond approved purposes, or other misuse of data; (b) the right of the data provider to

	external organisations, defining accountability	audit data flow and use; (c) obligations and accountability, including for improper use of data; (d) data retention and destruction (see also H24 below); (e) requirements relating to results arising, such as open-access publication, including attribution and acknowledgements; (f) management of any relevant intellectual property rights, pre-existing or arising; (g) payments, if applicable; (h) response to any data security and/or protocol breaches, including obligations and sanctions. Legal oversight should be obtained, including about enforceability if some parties are based in another country. A data transfer agreement would be required if data is to be moved, but this should be avoided; see **H13**.
H20	**Recognise potential for delays**	Recognise that governance processes are often slow. This may be undesirable for research but is a consequence of careful regulation. Do not over-promise speed to would-be researchers. Recognise also that research can be slow, and do not expect all projects to be completed rapidly. However, the duration of permitted access needs to be explicit at all times (with a process to request extensions if required).
H21	**Acknowledge the benefits of unexpected discovery (serendipity) and the potential for unexpected weaknesses in data**	Data may reveal unexpected findings or be of worse quality than expected. While it is important for statistical reasons that researchers distinguish between pre-planned and "post hoc" questions, oversight groups should be supportive both of unexpected findings (143) and the need to analyse data in slightly different ways because of limitations discovered in the data (e.g. some data being missing but a proxy measure being available). Organisations should support researchers to create protocols that are flexible enough to acknowledge this in advance (e.g. specifying the purpose of analysis rather than attempting to pre-specify every test), and should allow rapid light-touch updating of protocols via a formal amendment process (see **H11**), as long as the changes are within the spirit of the original project, do not increase key risks (such as of inadvertent re-identification), and are within the broader governance rules. But be careful not to allow "creep" to a project that would not have been approved as a "new" application.
H22	**Public register of projects and outputs**	Maintain and promote a public and accessible register of projects involving access to health data without explicit consent. Ideally, include also those with explicit consent. ► Be clear about what data is being accessed, how, and why. Make the full protocol, and details of PPI for the project, publicly accessible. Include a plain-language summary.► Be clear about who has access, including any industrial/commercial involvement. ► When possible, make contracts and data access agreements public. ► Add projects to the register as soon as they have been approved. ► Make information on any data breaches (protocol or security breaches) public, without disclosing details that might themselves compromise security. ► Add outputs (e.g. publications) to the register, to support public understanding and verification of the uses to which data is put.
H23	**Monitor publications and other outputs**	Require researchers to report outputs (e.g. publications) to the data provider. Audit compliance with the approved protocol. Add outputs to the public register (see above).
H24	**Monitor data retention/destruction**	Ensure that data is retained for long enough to be used and then to comply with data retention requirements. Then ensure that any copies (e.g. within a TRE) are destroyed securely when no longer required (ideally automatically). There should not be copies outside NHS environments anyway (see **H13/H18**). It is recognised that research data may be stored for many years, and that original NHS medical records have additional, and sometimes complex, retention rules (144). If disposing of digital storage equipment, recognise that simple deletion is inadequate and different from destruction or "sanitisation" (145).

Table 7: Guidelines for commercial (C) applicants seeking to use mental health data for research (potential data users). For abbreviations, see **Abbreviations**.

Category	*Number*	*Summary*	*Detail*
Project planning			
	C1	**Seek PPI input early**	Seek diverse patient and public involvement (PPI) before applying for data. Ideally, seek it at the stage of designing your research programme. There is now an NIHR service to assist commercial organisations with PPI (146). Consider including public co-applicants [e.g. (147)]. Ask potential collaborating NHS organisations for help if you are unfamiliar with this process.
	C2	**Define requirements clearly**	Specify clearly what data you want access to, and why (including the expected patient/public benefit). A plain-language summary, sufficiently detailed, must be included. (See H11, **Table 6**.)
	C3	**Train staff in information governance**	Ensure that all staff involved with research understand relevant laws, regulations, and local policies regarding the handling of sensitive health data, and know how to apply them in practice. Ensure staff understand the penalties for data mishandling. There should be mechanisms for keeping staff up to date and from learning from any breaches.
	C4	**Train staff in statistical disclosure control**	Ensure that research staff understand the methods required to produce safe outputs via statistical disclosure control (SDC). Consider national standard training, such as the Office for National Statistics (ONS) Accredited Researcher scheme (140).
	C5	**Distinguish data collected with consent from routinely collected data without research consent**	You may wish to collect data directly from participants with their explicit consent, and/or you may wish to use routinely collected health data (without consent, though likely subject to opt-outs). At all times, distinguish these two categories of data. They differ in sensitivity, and also in law (3,4).
	C6	**Be transparent**	Announce your intentions publicly as soon as possible. Ensure collaborating NHS organisations are involved in communications. Be prepared to respond to public challenge. Pass on relevant queries to participating NHS organisations.
	C7	**Understand when NHS REC approvals are required**	► All research involving NHS patients (people recruited because of their use of NHS services) directly require NHS Research Ethics Committee (REC) approval (53). ► Research databases may be reviewed by RECs (53). It is strongly recommended to seek REC approval for these. ► Some NHS organisations hold REC approval for research databases, and then further work using those databases may be permitted via local approvals, under the overarching REC approval. ► Market research of many kinds (meaning research for a commercial purpose) does not normally require NHS REC review (53), though other guidelines and oversight bodies exist (90,91). ► Many kinds of research involving employees do not require NHS REC review (53). ► However, there are also kinds of non-market research involving the general public that might involve MH data. For example, a company might publish a consumer app available to download from a public app store, and then seek to analyse MH data collected in this way. This might not be judged to require NHS REC review (148). In the context of a university, this sort of research would require institutional REC review. The fact that such research may be exempt from REC review is a potential risk and means it is particularly important for companies to consider how participants are informed, how consent is sought, and how data is handled in this research context. (See also **Table 8**.)
Application			
	C8	**Avoid secrecy**	You should not seek non-disclosure agreements (NDAs) either (a) from consenting participants providing their healthcare data, or (b) from healthcare organisations providing some form of access to routinely collected data without consent. This is part of transparency (see above). It may be acceptable to seek an NDA for other aspects (e.g. initial scoping) (133) but their use should be limited. If third-party organisations are used (e.g. for recruitment), participants should be made aware of all organisations involved.
	C9	**Define a clear protocol and plain-language summary**	Whether or not research needs external approvals, it is good practice to create a detailed protocol in advance. NHS RECs will require this. In some fields it might be encouraged to publish protocols in advance as part of study pre-registration. Protocols should include a plain-language summary that is sufficiently detailed. Check this is comprehensible by asking an independent non-expert to review it. (See H11, **Table 6**.)
	C10	**Seek approvals**	If applicable, seek the relevant research approvals. Expect some delays and the potential for revisions.
Execution			
	C11	**Follow the protocol carefully**	You should be diligent in following your protocol and contractual requirements (see **Table 6**), including any terms of approval, and in reporting data breaches if any occur. Do not use data for purposes beyond the approved protocol, or link data to other data sets outside your approvals (see also **C12** and **H11/H19, Table 6**). Make no attempt to re-identify data subjects.
	C12	**Request protocol and project amendments promptly**	Apply promptly to the relevant oversight body if new analyses or approaches are required. (See H11, **Table 6**.) Remember to update public-facing materials.
	C13	**Ensure quality of the research**	Check and report the quality of all aspect of the research, and its limitations. Examples are data quality, coverage and completeness, adequacy of analytical methods, potential for bias, and generalisability and clinical relevance of the results. Ensure that you have all the necessary expertise within your team, recognising that this may expand across multiple disciplines.

C14	**Ensure statistical disclosure control**	Apply recognised SDC methods (108,149) to create safe outputs for publication. Manual checks should be made even if automatic SDC is used. Comply with the host organisation's requirements and checks around SDC (see **Table 6**).
C15	**Report publications and other outputs**	Ensure that the results of your research are publicly available. Report them to participating health organisations. Ensure any consenting participants have the option to be sent the report. Provide a summary of the research in plain language, including all key characteristics and outputs of the study, the ways in which patients and the public were involved, and the way(s) in which health data was used. Link outputs to your public materials about the project, so the public can see the consequences and any benefits of the project.

Table 8: Recommendations for NHS (N) executive/regulatory bodies. Most do not relate to commercial/industrial research specifically but are more generic. For abbreviations, see **Abbreviations**.

Category	*Number*	*Summary*	*Detail*
Public choice			
	N1	**Make it easy for citizens to choose about NHS data sharing and research participation**	Provide a single UK-wide secure authenticated site (or 4×nation-specific sites), linked to national "Spine" identity services, for citizens to (a) choose whether to pre-consent for their data to be shared across the NHS for their personal direct clinical care; (b) choose the uses of their NHS data for research without specific consent, e.g. the current NHS National Data Opt-Out; and (c) choose whether to sign up to be contacted directly about participatory research, on an opt-in basis [e.g. SHARE in Scotland (150)]. There is strong public support for a web site for these (6).
Regulatory advice			
	N2	**Provide clear unified advice for NHS governance teams about service evaluation versus research, and about research database creation**	The HRA make a distinction between research and NHS audit/service evaluation. Research generally requires REC review, while audit/service evaluation is conducted by NHS bodies by themselves (151). This distinction is most difficult when routinely collected data is analysed. The distinction then largely centres around whether a study's conclusions are intended to apply to specific services, or to produce "generalisable" or "transferable" knowledge. If the latter, this is research (151,152), and usually (but not always) requires REC review (151). Other guidance states that "research limited to secondary use of information previously collected in the course of normal care (without an intention to use it for research at the time of collection) is generally excluded from REC review, provided that the patients or service users are not identifiable to the research team in carrying out the research" (153). Guidance to RECs indicates that applications for REC review of research databases are therefore normally voluntary (154), and separate HRA approvals are also not required (155). In our view, REC review adds independent oversight and the transparency associated with a central register of REC-reviewed projects. We strongly recommend that REC review should be required for all research databases using NHS patient data.
	N3	**Provide unified advice for NHS governance teams about when NHS Act Section 251 (or equivalent) approval is required**	The CAG offer general advice about when NHS Act section 251 (s251) approval is required, and some examples of precedent (156,157). However, the CAG require that data controllers (e.g. NHS Trusts) make the decision about whether s251 approval should be sought (i.e. whether data is identifiable) (156). Offering a formal regulatory decision to organisations that are in doubt may be helpful (e.g. confirmation that a proposed de-identified research database in a suitable TRE does not require s251 approval).
	N4	**Provide independent ethical review services for companies conducting non-NHS health-related research**	The HRA for England and Wales (158), equivalent organisations in the UK Devolved Administrations (**Supplementary Table 4**), and the UK-wide NHS Research Ethics Service (159) review research involving the NHS. This includes all drug trials, medical device research, most tissue research, all studies involving participants identified in connection with their or their families' use of NHS services, and some other categories (148). Outside this context, universities and other higher education institutions typically require research projects to pass review by an internal research ethics committee if they involve human participants (148,160). However, we are unaware of any process for companies to seek independent ethical review of research involving (for example) consenting participants providing health-related data directly to a company. We suggest that such a service be made available to industry for health-related research, on a voluntary basis with cost recovery, with transparent processes as below. It should incorporate PPI input.
Privacy			
	N5	**Support de-identification and de-identified linkage whenever practical**	De-identification is central to health data research without explicit patient consent. It should be legally permitted that NHS organisations can, as Data Controllers, instruct a trusted Data Processor to perform de-identification for research, encouraging de-identification at the earliest possible opportunity. NHS central bodies sometimes provide data as part of linkage studies [e.g. (161)]. In principle, data from multiple sources can be linked identifiably or without direct identifiers. Identifiable linkage of UK health data is likely to need s251 support (**Supplementary Table 3**) for the use of confidential patient information for research without consent, even if identifiers are removed before researchers are given access. This route is intended as a last resort when no other methods are feasible (54,156). At present, some NHS bodies reject requests for de-identified linkage on grounds of effort/impracticality, requiring instead identifiable linkage [personal communication, February 2023]. However, de-identified linkage is quick and straightforward using encrypted person-unique identifiers, and feasible in more complex situations (67). Some studies will continue to require identifiable linkage (e.g. studies requiring linkage verification), but we recommend that de-identified linkage be supported whenever practical, to reduce the use of identifiable information.
Transparency			
	N6	**Continue to support citizens to view their own NHS data, and for NHS organisations to implement legally required safeguards**	Some UK citizens have access to part of their medical records via central systems like the NHS App (162). At present this is for primary care (GP) records, including information sent from secondary care (e.g. hospitals) to GP surgeries. Some individual EHR systems offer further "patient portals", including for some secondary care data. Patients should have access to their full medical records in most cases. Additionally, it should be easy, in all EHR systems, for staff to mark aspects of records according to the exclusions currently required—if rarely—from self-access: (a) information relating to an identifiable third party (excluding staff) where that third party hasn't given consent, or (b) information that might cause serious harm to any individual (3,163). ["Live" online self-access requires that all records be checked by the NHS against these exclusions at the time of creation, rather than checking a small number of records upon request (164).] The NHS should require that all EHR systems make this form of marking easy, to maintain safety and thus support broader patient access to records. However, digital access must not become the only means for patients to access their records, to avoid "digital exclusion".
	N7	**Make public**	Summary information written in plain language should be available and readily searchable for all studies

		registers of studies approved by NHS RECs and other oversight bodies accessible and readily searchable	approved by an NHS REC. At present, summaries of projects considered by NHS RECs are available (165) but are not indexed by public search engines, and so are hard to find. Summary information is already available for CAG approvals (166); this is indexed by search engines but is not in an accessible web format. Summaries should make any commercial/industrial role clear, where that role is not simply as a data processor.
	N8	**Unify and simplify research reporting requirements**	There are multiple reporting requirements for researchers. A centralised simple reporting system with a good public interface, including an application programming interface (API), would aid this. Such a system could index and link research projects, their PPI, plain-language summaries, institutions, funding, and publications.
Support			
	N9	**Endorse guidelines for data-controlling organisations and applicants**	Central bodies should endorse patient/public-supported guidelines for data-controlling organisations and data applicants, such as those provided in **Table 6** and **Table 7**.

10 REFERENCES

1. NHS England. The NHS Constitution for England. (2013) https://www.gov.uk/government/publications/the-nhs-constitution-for-england

2. NHS England. The NHS Constitution for England. (2015) https://www.gov.uk/government/publications/the-nhs-constitution-for-england

3. UK. Data Protection Act 2018. (2018). https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted

4. European Parliament and Council. Regulation (EU) 2016/679 (General Data Protection Regulation). *Official Journal of the European Union* (2016) L119:1–88.

5. NHS Health Research Authority. Patient information and health and care research. (2021) https://web.archive.org/web/20230331215604/https://www.hra.nhs.uk/information-about-patients/

6. Jones LA, Nelder JR, Fryer JM, Alsop PH, Geary MR, Prince M, Cardinal RN. Public opinion on sharing data from health services for clinical and research purposes without explicit consent: an anonymous online survey in the UK. *BMJ Open* (2022) 12:e057579. doi: 10.1136/bmjopen-2021-057579

7. Zhang J, Morley J, Gallifant J, Oddy C, Teo JT, Ashrafian H, Delaney B, Darzi A. Mapping and evaluating national data flows: transparency, privacy, and guiding infrastructural transformation. *Lancet Digit Health* (2023) 5:e737–e748. doi: 10.1016/S2589-7500(23)00157-7

8. UK Department for Business, Energy & Industrial Strategy. Life Sciences Vision (policy paper). (2021) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1000030/life-sciences-vision.pdf

9. Donelan M, The Viscount Camrose. Data Protection and Digital Information Bill. (2024). https://bills.parliament.uk/bills/3430 [Accessed July 2, 2024]

10. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. *J Med Ethics* (2015) 41:404–409. doi: 10.1136/medethics-2014-102374

11. Ghafur S, Van Dael J, Leis M, Darzi A, Sheikh A. Public perceptions on data sharing: key insights from the UK and the USA. *Lancet Digit Health* (2020) 2:e444–e446. doi: 10.1016/S2589-7500(20)30161-8

12. Understanding Patient Data, Ada Lovelace Institute. Foundations of fairness: where next for NHS health data partnerships? (2020). https://understandingpatientdata.org.uk/sites/default/files/2020-03/Foundations%20of%20Fairness%20-%20Summary%20and%20Analysis.pdf [Accessed April 20, 2020]

13. Harkness F, Rijnevald C, Liu Y, Kashef S, Cowan M. A UK-wide public dialogue exploring what the public perceive as 'public good' use of data for research and statistics. (2022) https://www.adruk.org/fileadmin/uploads/adruk/Documents/PE_reports_and_documents/ADR_UK_OSR_Public_Dialogue_final_report_October_2022.pdf

14. Understanding Patient Data. Putting people at the centre of decisions about patient data. (2023) https://understandingpatientdata.org.uk/ [Accessed May 3, 2023]

15. Smyth C. Rishi Sunak urged to let drug firms use NHS data for new medicines. *The Times* (2023) https://www.thetimes.co.uk/article/rishi-sunak-nhs-data-drugs-medicine-2023-rdg7ln3p5

16. Serle J. Government to put 'realistic' price on NHS data. *Health Service Journal* https://www.hsj.co.uk/finance-and-efficiency/government-to-put-realistic-price-on-nhs-data/7040105.article [Accessed October 1, 2025]

17. Wellcome Trust. The one-way mirror: Public attitudes to commercial access to health data. (2016) https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

18. Critical Illness Definitions and Geographical Variations Working Party. Extending the critical path. (2013) https://sias.org.uk/media/1161/extending-the-critical-path.pdf

19. Avery MC. Supporting NHS hospitals in England to engage with acceptable and ethical use of health and care data by commercial organisations. Cambridge, UK: University of Cambridge. (2021).

20. Richards SJ. Applying Survival Models to Pensioner Mortality Data. *British Actuarial Journal* (2008) 14:257–303. doi: 10.1017/S1357321700001720

21. Harkness F, Blodgett J, Rijneveld C, Waind E, Amugi M, McDonald F. Building a trustworthy national data research infrastructure: A UK-wide public dialogue. DARE UK, ADR UK, HDR UK, UKRI, Kohlrabi Consulting. (2022). https://dareuk.org.uk/wp-content/uploads/2022/05/DARE_UK_Building_a_Trustworthy_National_Data_Research_Infrastructure_Public_Dialogue_May-2022.pdf

22. Health Data Research UK. Access to health data for research and innovation – frequently asked questions. (2020) https://www.hdruk.ac.uk/wp-content/uploads/2020/01/Using-health-data-for-research-questions-for-public-January-2020-1.pdf [Accessed February 14, 2023]

23. Health Data Research UK, Academic of Medical Sciences, Collaboration for the Advancement of Sustainable Medical Innovation. Realising patient and NHS benefits from health and care data – from policy to practice. (2020) https://www.hdruk.ac.uk/wp-content/uploads/2020/01/200227-Realising-patient-and-NHS-benefits-from-health-and-care-data-from-policy-to-practice.pdf [Accessed February 14, 2023]

24. Dowden A. Data Access Processes: Public Involvement, Trust and Transparency. (2021) https://www.hdruk.ac.uk/wp-content/uploads/2021/06/UPDATED-Paper-C-210414-Full-Report-Data-Access-Processes-Survey.pdf

25. Health Data Research UK. Health Data Research Hubs: Legal Materials and Templates: Master Services Agreement (template and guidance note). (2021) https://www.hdruk.ac.uk/about-us/policies/health-data-research-hubs-legal-materials-and-templates/ [Accessed February 14, 2023]

26. Health Data Research UK. Case study: Data-Driven Commercial Services - BREATHE hub. (2022) https://www.hdruk.ac.uk/case-studies/data-driven-commercial-services-breathe-hub/

27. Department for Health and Social Care, Hancock M. Coronavirus (COVID-19): notification to organisations to share information. (2020) https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information

28. Sterckx S, Rakic V, Cockbain J, Borry P. "You hoped we would sleep walk into accepting the collection of our data": controversies surrounding the UK care.data scheme and their wider relevance for biomedical research. *Med Health Care Philos* (2015) doi: 10.1007/s11019-015-9661-6

29. Coulter A. Patient trust in plans to share primary care data. *BMJ* (2021) 373:n1413. doi: 10.1136/bmj.n1413

30. Gayle D. GPs urged to refuse to hand over patient details. *The Guardian* (2021) https://www.theguardian.com/society/2021/jun/01/gps-urged-to-refuse-to-hand-over-patient-details-to-nhs-digital [Accessed June 29, 2021]

31. Crouch H. GPDPR September implementation date is scrapped. *Digital Health* (2021) https://www.digitalhealth.net/2021/07/gpdpr-september-implementation-date-scrapped/ [Accessed January 27, 2022]

32. Understanding Patient Data. Annual report 2016-2017. (2017) https://understandingpatientdata.org.uk/sites/default/files/2017-12/UPD%20Annual%20Report%202016-2017.pdf

33. Goldacre B, Morley J. Better, Broader, Safer: Using health data for research and analysis. A review commissioned by the Secretary of State for Health and Social Care. (2022) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1067053/goldacre-review-using-health-data-for-research-and-analysis.pdf

34. Hoeksma J. First four regional secure data environments confirmed. *Digital Health* (2022) https://www.digitalhealth.net/2022/12/first-four-regional-secure-data-environments-confirmed/

35. NHS England. Digitising, connecting and transforming health and care. (2023) https://web.archive.org/web/20230421151841/https://www.england.nhs.uk/digitaltechnology/digitising-connecting-and-transforming-health-and-care/

36. DARE UK Delivery Team. Federated architecture blueprint. (2023) https://dareuk.org.uk/wp-content/uploads/2023/04/DARE-UK-Federated-Architecture-1-Initial.pdf [Accessed June 14, 2023]

37. Faculty of Clinical Informatics. NHS Federated Data Platform: FCI Position Paper. (2023) https://web.archive.org/web/20230706110229/https://fci.org.uk/static/888f2a69-71d1-497e-a5f9e4959cda8a69/FCIPositionPaperFederatedDataPlatform.pdf

38. Delacroix S, Lawrence ND. Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law* (2019) 9:236–252. doi: 10.1093/idpl/ipz014

39. Greater Manchester Combined Authority. Case study: Can a data trust give people the control they want over how their data is shared? (2022) https://web.archive.org/web/20230330054419/https://www.greatermanchester-ca.gov.uk/what-we-do/digital/case-studies/case-study-can-a-data-trust-give-people-the-control-they-want-over-how-their-data-is-shared/

40. Information Governance Services. Data Saves Lives and the NHS: Legal and Ethical Considerations to Guide the NHS in its Digital Transformation. (2022) https://www.informationgovernanceservices.com/wp-content/uploads/2023/06/Data-Saves-Lives-and-the-NHS-Legal-and-ethical-considerations-to-guide-the-NHS-in-its-digital-transformation.pdf [Accessed June 19, 2023]

41. Kirkham EJ, Lawrie SM, Crompton CJ, Iveson MH, Jenkins ND, Goerdten J, Beange I, Chan SWY, McIntosh A, Fletcher-Watson S. Experience of clinical services shapes attitudes to mental health data sharing: findings from a UK-wide survey. *BMC Public Health* (2022) 22:357. doi: 10.1186/s12889-022-12694-z

42. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, Koschorke M, Shidhaye R, O'Reilly C, Henderson C. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. *Lancet* (2016) 387:1123–1132. doi: 10.1016/S0140-6736(15)00298-6

43. Woelbert E, White R, Lundell-Smith K, Grant J, Kemmer D. Inequities of mental health research funding. (2020) https://www.drugsandalcohol.ie/33429/1/The_Inequities_of_Mental_Health_Research.pdf

44. Kirkham EJ, Iveson M, Beange I, Crompton CJ, McIntosh A, Fletcher-Watson S. A stakeholder-derived best practice checklist for mental health data science in the UK. (2020) https://mhdss.ac.uk/best-practice-mental-health-data-science

45. Kirkham EJ, Crompton CJ, Iveson MH, Beange I, McIntosh AM, Fletcher-Watson S. Co-development of a Best Practice Checklist for Mental Health Data Science: A Delphi Study. *Front Psychiatry* (2021) 12:643914. doi: 10.3389/fpsyt.2021.643914

46. UK. Equality Act 2010. (2010) https://www.legislation.gov.uk/ukpga/2010/15/contents [Accessed July 5, 2021]

47. National Institute for Health and Care Research. Payment guidance for researchers and professionals. (2022) https://www.nihr.ac.uk/documents/payment-guidance-for-researchers-and-professionals/27392

48. NIHR School for Primary Care Research. Do I need ethical approval to run an involvement activity? (2023) https://www.spcr.nihr.ac.uk/PPI/resources-for-researchers/faq/do-i-need-ethical-approval-to-run-an-involvement-activity [Accessed March 15, 2023]

49. World Medical Association. Declaration of Helsinki - ethical principles for medical research involving human subjects. (2013) https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [Accessed March 22, 2021]

50. UK. Mental Capacity Act 2005. (2005). https://www.legislation.gov.uk/ukpga/2005/9/contents

51. Dayé C. How to train your oracle: The Delphi method and its turbulent youth in operations research and the policy sciences. *Soc Stud Sci* (2018) 48:846–868. doi: 10.1177/0306312718798497

52. DATAMIND. An industry guide to accessing and using mental health data. (2023) https://web.archive.org/web/20231109140849/https://datamind.org.uk/wp-content/uploads/2023/10/Industry-Guide.pdf

53. NHS Health Research Authority. Governance arrangements for Research Ethics Committees. (2021) https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/governance-arrangement-research-ethics-committees/ [Accessed March 15, 2023]

54. Cross L, Carson LE, Jewell A, Heslin M, Osborn D, Downs J, Stewart R. Guidance for researchers wanting to link NHS data using non-consent approaches: a thematic analysis of feedback from the Health Research Authority Confidentiality Advisory Group. *Int J Popul Data Sci* (2020) 5:1355. doi: 10.23889/ijpds.v5i1.1355

55. Stewart R, Soremekun M, Perera G, Broadbent M, Callard F, Denis M, Hotopf M, Thornicroft G, Lovestone S. The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. *BMC Psychiatry* (2009) 9:51. doi: 10.1186/1471-244X-9-51

56. Sieberts SK, Marten C, Bampton E, Björling EA, Burn A-M, Carey EG, Carlson S, Fernandes B, Kalha J, Lindani S, et al. Young people's data governance preferences for their mental health data: MindKind Study findings from India, South Africa, and the United Kingdom. *PLoS One* (2023) 18:e0279857. doi: 10.1371/journal.pone.0279857

57. Sudlow CLM. Uniting the UK's health data: a huge opportunity for society. (2024) https://doi.org/10.5281/zenodo.13353747

58. Mader LB, Harris T, Kläger S, Wilkinson IB, Hiemstra TF. Inverting the patient involvement paradigm: defining patient led research. *Res Involv Engagem* (2018) 4:21. doi: 10.1186/s40900-018-0104-4

59. Macaulay Y, UK Health Data Research Alliance, Pan-UK Data Governance Steering Group, HDR UK Public Advisory Board, Five Safes Action Force. Pan-UK Data Governance Steering Group Data Access Transparency Standards. (2023) doi: 10.5281/zenodo.8262453

60. Price A, Farooq R, Yuan J-M, Menon VB, Cardinal RN, O'Brien JT. Mortality in dementia with Lewy bodies compared with Alzheimer's dementia: a retrospective naturalistic cohort study. *BMJ Open* (2017) 7:e017504. doi: 10.1136/bmjopen-2017-017504

61. Ford E, Pillinger S, Stewart R, Jones K, Roberts A, Casey A, Goddard K, Nenadic G. What is the patient re-identification risk from using de-identified clinical free text data for health research? *AI Ethics* (2025) doi: 10.1007/s43681-025-00681-0

62. Castell S, Robinson L, Ashford H. Future data-driven technologies and the implications for use of patient data: dialogue with public, patients and healthcare professionals. (2018) https://acmedsci.ac.uk/file-download/6616969

63. Academy of Medical Sciences. Our data-driven future in healthcare: people and partnerships at the heart of health related technologies. (2018) https://acmedsci.ac.uk/file-download/74634438

64. Jones KH, Ford DV, Lyons RA. The SAIL Databank: 10 years of spearheading data privacy and research utility, 2007-2017. (2017) doi: http://dx.doi.org/10.23889/

65. Fernandes AC, Cloete D, Broadbent MTM, Hayes RD, Chang C-K, Jackson RG, Roberts A, Tsang J, Soncul M, Liebscher J, et al. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. *BMC Med Inform Decis Mak* (2013) 13:71. doi: 10.1186/1472-6947-13-71

66. Cardinal RN. Clinical records anonymisation and text extraction (CRATE): an open-source software system. *BMC Med Inform Decis Mak* (2017) 17:50. doi: 10.1186/s12911-017-0437-1

67. Cardinal RN, Moore A, Burchell M, Lewis JR. De-identified Bayesian personal identity matching for privacy-preserving record linkage despite errors: development and validation. *BMC Medical Informatics and Decision Making* (2023) 23:85. doi: 10.1186/s12911-023-02176-6

68. NHS Health Research Authority. Confidentiality Advisory Group. (2019) https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/confidentiality-advisory-group/ [Accessed July 23, 2019]

69. Information Commissioner's Office. Information Commissioner's Office. (2023) https://ico.org.uk/

70. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, Staplin N, Brightling C, Ustianowski A, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med* (2021) 384:693–704. doi: 10.1056/NEJMoa2021436

71. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annu Rev Public Health* (2016) 37:61–81. doi: 10.1146/annurev-publhealth-032315-021353

72. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* (2020) 584:430–436. doi: 10.1038/s41586-020-2521-4

73. Joint Committee on Vaccination and Immunisation. *Advice on priority groups for COVID-19 vaccination*. (2020). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/940396/Priority_groups_for_coronavirus__COVID-19__vaccination_-_advice_from_the_JCVI__2_December_2020.pdf

74. National Audit Office. The rollout of the COVID-19 vaccination programme in England (Department of Health & Social Care and Department for Business, Energy & Industrial Strategy). (2022) https://www.nao.org.uk/wp-content/uploads/2022/02/The-rollout-of-the-COVID-19-vaccination-programme-in-England.pdf

75. Taquet M, Luciano S, Geddes JR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. *The Lancet Psychiatry* (2020) doi: 10.1016/S2215-0366(20)30462-4

76. Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, White IR, Caulfield MJ, Deanfield JE, Smeeth L, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. *Lancet* (2014) 383:1899–1911. doi: 10.1016/S0140-6736(14)60685-1

77. Maas J, Verheij RA, de Vries S, Spreeuwenberg P, Schellevis FG, Groenewegen PP. Morbidity is related to a green living environment. *J Epidemiol Community Health* (2009) 63:967–973. doi: 10.1136/jech.2008.079038

78. Osborn DP, Hardoon S, Omar RZ, Holt RI, King M, Larsen J, Marston L, Morris RW, Nazareth I, Walters K, et al. The PRIMROSE cardiovascular risk prediction models for people with severe mental illness: Results from the PRIMROSE Research Programme: PRedIction and Management of cardiovascular Risk in peOple with SEvere mental illnesses. *JAMA psychiatry* (2015) 72:143. doi: 10.1001/jamapsychiatry.2014.2133

79. Ray WA, Chung CP, Murray KT, Hall K, Stein CM. Atypical antipsychotic drugs and the risk of sudden cardiac death. *N Engl J Med* (2009) 360:225–235. doi: 10.1056/NEJMoa0806994

80. Scherrer JF, Garfield LD, Chrusciel T, Hauptman PJ, Carney RM, Freedland KE, Owen R, True WR, Lustman PJ. Increased risk of myocardial infarction in depressed patients with type 2 diabetes. *Diabetes Care* (2011) 34:1729–1734. doi: 10.2337/dc11-0031

81. Grippo AJ, Johnson AK. Biological mechanisms in the relationship between depression and heart disease. *Neurosci Biobehav Rev* (2002) 26:941–962. doi: 10.1016/s0149-7634(03)00003-4

82. Velosa J, Delgado A, Finger E, Berk M, Kapczinski F, de Azevedo Cardoso T. Risk of dementia in bipolar disorder and the interplay of lithium: a systematic review and meta-analyses. *Acta Psychiatr Scand* (2020) doi: 10.1111/acps.13153

83. Chen S, Underwood BR, Jones PB, Lewis JR, Cardinal RN. Association between lithium use and the incidence of dementia and its subtypes: A retrospective cohort study. *PLoS Med* (2022) 19:e1003941. doi: 10.1371/journal.pmed.1003941

84. Forlenza OV, Radanovic M, Talib LL, Gattaz WF. Clinical and biological effects of long-term lithium treatment in older adults with amnestic mild cognitive impairment: randomised clinical trial. *Br J Psychiatry* (2019) 215:668–674. doi: 10.1192/bjp.2019.76

85. Eyting M, Xie M, Michalik F, Heß S, Chung S, Geldsetzer P. A natural experiment on the effect of herpes zoster vaccination on dementia. *Nature* (2025) doi: 10.1038/s41586-025-08800-x

86. O'Brien EC, Raman SR, Ellis A, Hammill BG, Berdan LG, Rorick T, Janmohamed S, Lampron Z, Hernandez AF, Curtis LH. The use of electronic health records for recruitment in clinical trials: a mixed methods analysis of the Harmony Outcomes Electronic Health Record Ancillary Study. *Trials* (2021) 22:465. doi: 10.1186/s13063-021-05397-0

87. Richards DA, Hill JJ, Gask L, Lovell K, Chew-Graham C, Bower P, Cape J, Pilling S, Araya R, Kessler D, et al. Clinical effectiveness of collaborative care for depression in UK primary care (CADET): cluster randomised controlled trial. *BMJ* (2013) 347:f4913. doi: 10.1136/bmj.f4913

88. Callard F, Broadbent M, Denis M, Hotopf M, Soncul M, Wykes T, Lovestone S, Stewart R. Developing a new model for patient recruitment in mental health services: a cohort study using Electronic Health Records. *BMJ Open* (2014) 4:e005654. doi: 10.1136/bmjopen-2014-005654

89. Tyrer P, Cooper S, Salkovskis P, Tyrer H, Crawford M, Byford S, Dupont S, Finnis S, Green J, McLaren E, et al. Clinical and cost-effectiveness of cognitive behaviour therapy for health anxiety in medical patients: a multicentre randomised controlled trial. *Lancet* (2014) 383:219–225. doi: 10.1016/S0140-6736(13)61905-4

90. Market Research Society (MRS). Code of Conduct. (2019) https://web.archive.org/web/20221005172956/https://www.mrs.org.uk/pdf/MRS-Code-of-Conduct-2019.pdf

91. British Healthcare Business Intelligence Association (BHBIA). Legal and Ethical Guidelines for Healthcare Market Research. (2023) https://web.archive.org/web/20230307164746/https://www.bhbia.org.uk/assets/Downloads/Guidelines/bhbia_legal_and_ethical_guidelines_jan2023_final_inc_practicalexamples_locked2.pdf

92. Kendall-Raynor P. An Epic outcome. *Emergency Nurse* (2018) 26:12. doi: 10.7748/en.26.2.12.s10

93. Care Quality Commission. Cambridge University Hospitals NHS Foundation Trust: Inspection Report. (2019) https://www.cqc.org.uk/sites/default/files/new_reports/AAAJ0418.pdf

94. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, Pestrue J, Phillips M, Konye J, Penoza C, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med* (2021) 181:1065–1070. doi: 10.1001/jamainternmed.2021.2626

95. Makam AN, Nguyen OK, Auerbach AD. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: A systematic review. *J Hosp Med* (2015) 10:396–402. doi: 10.1002/jhm.2347

96. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. *Clin Radiol* (2019) 74:357–366. doi: 10.1016/j.crad.2019.02.006

97. Guo T, Xiang Y-T, Xiao L, Hu C-Q, Chiu HFK, Ungvari GS, Correll CU, Lai KYC, Feng L, Geng Y, et al. Measurement-Based Care Versus Standard Care for Major Depression: A Randomized Controlled Trial With Blind Raters. *Am J Psychiatry* (2015) 172:1004–1013. doi: 10.1176/appi.ajp.2015.14050652

98. Catarino A, Bateup S, Tablan V, Innes K, Freer S, Richards A, Stott R, Hollon SD, Chamberlain SR, Hayes A, et al. Demographic and clinical predictors of response to internet-enabled cognitive-behavioural therapy for depression and

anxiety. *BJPsych Open* (2018) 4:411–418. doi: 10.1192/bjo.2018.57

99. Ewbank MP, Cummins R, Tablan V, Bateup S, Catarino A, Martin AJ, Blackwell AD. Quantifying the Association Between Psychotherapy Content and Clinical Outcomes Using Deep Learning. *JAMA Psychiatry* (2020) 77:35–43. doi: 10.1001/jamapsychiatry.2019.2664

100. Catarino A, Fawcett JM, Ewbank MP, Bateup S, Cummins R, Tablan V, Blackwell AD. Refining our understanding of depressive states and state transitions in response to cognitive behavioural therapy using latent Markov modelling. *Psychol Med* (2022) 52:332–341. doi: 10.1017/S0033291720002032

101. Ewbank MP, Cummins R, Tablan V, Catarino A, Buchholz S, Blackwell AD. Understanding the relationship between patient language and outcomes in internet-enabled cognitive behavioural therapy: A deep learning approach to automatic coding of session transcripts. *Psychother Res* (2021) 31:326–338. doi: 10.1080/10503307.2020.1788740

102. Rogers P, Wang D, Lu Z. Medical Information Mart for Intensive Care: A Foundation for the Fusion of Artificial Intelligence and Real-World Data. *Front Artif Intell* (2021) 4:691626. doi: 10.3389/frai.2021.691626

103. Johnson AEW, Pollard TJ, Shen L, Lehman L-WH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. *Sci Data* (2016) 3:160035. doi: 10.1038/sdata.2016.35

104. Information Commissioner's Office. Chapter 1: introduction to anonymisation [Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance]. (2021) https://web.archive.org/web/20211109073338/https://ico.org.uk/media/about-the-ico/consultations/2619862/anonymisation-intro-and-first-chapter.pdf [Accessed February 13, 2022]

105. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. *PLoS One* (2011) 6:e28071. doi: 10.1371/journal.pone.0028071

106. Culnane C, Rubinstein BIP, Teague V. Health Data in an Open World. *arXiv:171205627 [cs]* (2017) http://arxiv.org/abs/1712.05627

107. Information Commissioner's Office. Chapter 2: How do we ensure anonymisation is effective? [Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance]. (2021) https://web.archive.org/web/20220124123150/https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf [Accessed February 13, 2022]

108. Griffiths E, Greci C, Kotrotsios Y, Parker S, Scott J, Welpton R, Wolters A, Woods C. Handbook on Statistical Disclosure Control for Outputs. (2019) https://web.archive.org/web/20201026152627/https://ukdataservice.ac.uk/media/622521/thf_datareport_aw_web.pdf

109. Information Commissioner's Office. Enforcement action. (2023) https://ico.org.uk/action-weve-taken/enforcement/ [Accessed April 25, 2023]

110. Hoeksma J. Client data exfiltrated in Advanced NHS cyber attack. *Digital Health* (2022) https://www.digitalhealth.net/2022/10/client-data-exfiltrated-advanced-nhs-cyber-attack/

111. Department for Science, Innovation and Technology, Donelan M. British Businesses to Save Billions Under New UK Version of GDPR. *GOVUK* (2023) https://www.gov.uk/government/news/british-businesses-to-save-billions-under-new-uk-version-of-gdpr [Accessed March 18, 2023]

112. Smith J. Patient confidentiality and consent to publication. *BMJ* (2008) 337:a1572. doi: 10.1136/bmj.a1572

113. Shelby A, Ernst K. Story and science: how providers and parents can utilize storytelling to combat anti-vaccine misinformation. *Hum Vaccin Immunother* (2013) 9:1795–1801. doi: 10.4161/hv.24828

114. Koslap-Petraco M. Vaccine hesitancy: Not a new phenomenon, but a new threat. *J Am Assoc Nurse Pract* (2019) 31:624–626. doi: 10.1097/JXX.0000000000000342

115. Flaherty DK. The vaccine-autism connection: a public health crisis caused by unethical medical practices and fraudulent science. *Ann Pharmacother* (2011) 45:1302–1304. doi: 10.1345/aph.1Q318

116. Jayanetti C. NHS data grab on hold as millions opt out. *The Guardian* (2021) http://www.theguardian.com/society/2021/aug/22/nhs-data-grab-on-hold-as-millions-opt-out [Accessed August 22, 2021]

117. Burr C. Charities are contributing to growing mistrust of mental-health text support — here's why. *The Conversation* (2022) http://theconversation.com/charities-are-contributing-to-growing-mistrust-of-mental-health-text-support-heres-why-179056 [Accessed March 13, 2023]

118. Gilburt H, Mallorie S. Mental health 360: quality and patient experience. (2024) https://www.kingsfund.org.uk/insight-and-analysis/long-reads/mental-health-360

119. British Medical Association. "It's broken": Doctors' experiences on the frontline of a failing mental healthcare system. (2024) https://www.bma.org.uk/media/ddclsiii/bma-mental-health-report-2024-web-final.pdf

120. UK. The Data Protection (Adequacy) (United States of America) Regulations 2023. (2023). https://www.legislation.gov.uk/uksi/2023/1028/contents/made

121. Department for Science, Innovation and Technology. UK-US data bridge: explainer. (2023) https://www.gov.uk/government/publications/uk-us-data-bridge-supporting-documents/uk-us-data-bridge-explainer [Accessed May 1, 2025]

122. Hern A. Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind. *The Guardian* (2017) https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act [Accessed November 6, 2020]

123. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. *Health Technol* (2017)1–17. doi: 10.1007/s12553-017-0179-1

124. Information Commissioner's Office. RFA0627721 – provision of patient data to DeepMind. (2017) https://ico.org.uk/media/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf

125. Prismall v Google. (2023). https://www.judiciary.uk/wp-content/uploads/2023/05/Prismall-v-Google.pdf [Accessed June 14, 2023]

126. Thomas R. More than a million NHS patients' details compromised after cyberattack. *The Independent* (2023) https://www.independent.co.uk/news/health/nhs-patient-data-attack-b2364202.html

127. UK Research and Innovation. UKRI open access policy. (2021) https://www.ukri.org/publications/ukri-open-access-policy/

128. Wellcome Trust. Open access policy. (2021) https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy [Accessed May 2, 2023]

129. National Institute for Health and Care Research (NIHR). NIHR Open Access publication policy - for publications submitted on or after 1 June 2022. (2021) https://www.nihr.ac.uk/documents/nihr-open-access-publication-policy-for-publications-submitted-on-or-after-1-june-2022/28999

130. Downs JM, Ford T, Stewart R, Epstein S, Shetty H, Little R, Jewell A, Broadbent M, Deighton J, Mostafa T, et al. An approach to linking education, social care and electronic health records for children and young people in South London: a linkage study of child and adolescent mental health service data. *BMJ Open* (2019) 9:e024355. doi: 10.1136/bmjopen-2018-024355

131. de Man Y, Wieland-Jorna Y, Torensma B, de Wit K, Francke AL, Oosterveld-Vlug MG, Verheij RA. Opt-In and Opt-Out Consent Procedures for the Reuse of Routinely Recorded Health Data in Scientific Research and Their Consequences for Consent Rate and Consent Bias: Systematic Review. *Journal of Medical Internet Research* (2023) 25:e42131. doi: 10.2196/42131

132. Hossain A, Hannan A. Empowering underserved groups through access to electronic health records. *BMJ* (2023) 380:247. doi: 10.1136/bmj.p247

133. National Institute for Health and Care Research. NIHR Clinical Research Network: Confidential information arrangements for the Life Sciences Industry feasibility services. (2019) https://www.nihr.ac.uk/documents/crn-confidential-disclosure-agreement-process-summary/11451 [Accessed March 12, 2023]

134. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* (2007) 4:e297. doi: 10.1371/journal.pmed.0040297

135. Health Data Research UK Public Advisory Board. Building trust in data access through public involvement in governance. (2021) https://web.archive.org/web/20220804032353/https://www.hdruk.ac.uk/wp-content/uploads/2021/07/280621-PAB-Data-Access-procedures-paper-Building-trust-in-data-access-through-public-involvement-in-governance.pdf

136. Ritchie F. Secure access to confidential microdata: four years of the Virtual Microdata Laboratory. *Econ Lab Market Rev* (2008) 2:29–34. doi: 10.1057/elmr.2008.73

137. Office for National Statistics. What is the Five Safes framework? (2021) https://web.archive.org/web/20210902123414/https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/

138. UK Government. Setting up a social enterprise. (2023) https://www.gov.uk/set-up-a-social-enterprise [Accessed July 29, 2024]

139. Office for National Statistics. Access the data securely. (2024) https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/accessthedatasecurely [Accessed October 8, 2024]

140. UK. Digital Economy Act 2017. Queen's Printer of Acts of Parliament. (2017). https://www.legislation.gov.uk/ukpga/2017/30/contents/enacted [Accessed October 27, 2020]

141. Office for National Statistics. Become an accredited researcher. (2023) https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/becomeanaccreditedresearcher [Accessed March 13, 2023]

142. SafePod Network. The SafePod Network. (2025) https://safepodnetwork.ac.uk/ [Accessed July 9, 2025]

143. McCay-Peet L, Toms EG. Investigating serendipity: How it unfolds and what may influence it. *Journal of the Association for Information Science and Technology* (2015) 66:1463–1476. doi: 10.1002/asi.23273

144. NHS England. Records Management Code of Practice 2021. (2021) https://web.archive.org/web/20230316102300/https://transform.england.nhs.uk/media/documents/NHSX_Records_Management_CoP_V7.pdf

145. National Cyber Security Centre (NCSC). Secure sanitisation of storage media. (2020) https://web.archive.org/web/20221205150835/https://www.ncsc.gov.uk/guidance/secure-sanitisation-storage-media

146. National Institute for Health and Care Research (NIHR). Engage with patients and the public to improve your research design. (2024) https://www.nihr.ac.uk/partners-and-industry/industry/support-services-facilities/patient-engagement-service.htm [Accessed July 2, 2024]

147. National Institute for Health and Care Research (NIHR). Public co-applicants in research: guidance on roles and responsibilities. (2024) https://web.archive.org/web/20240715144833/https://www.learningforinvolvement.org.uk/content/resource/public-co-applicants-in-research-guidance-on-roles-and-responsibilities/ [Accessed October 8, 2024]

148. UK Medical Research Council, NHS Health Research Authority. Do I need NHS REC review? (2023) https://www.hra-decisiontools.org.uk/ethics/ [Accessed March 6, 2023]

149. Elliot M, Domingo-Ferrer J. “The future of statistical disclosure control.,” In: UK Government Data Quality Hub, editor. *Privacy and data confidentiality methods: a Data and Analysis Method Review (DAMR)*. UK Government Statistical Service (2018) http://arxiv.org/abs/1812.09204 [Accessed April 12, 2021]

150. NHS Research Scotland. What is SHARE? (2013) https://www.registerforshare.org/ [Accessed July 2, 2024]

151. NHS Health Research Authority. Is my study research? (2022) http://www.hra-decisiontools.org.uk/research/

152. Health and Social Care, NHS Research Scotland, Health and Care Research Wales, NHS Health Research Authority. UK Policy Framework for Health and Social Care Research. (2017) https://web.archive.org/web/20230129163958/https://s3.eu-west-2.amazonaws.com/www.hra.nhs.uk/media/documents/Final_Accessibility_uk-policy-framework-health-social-care-research_.pdf [Accessed March 6, 2023]

153. Health and Social Care, Health and Care Research Wales, NHS Research Scotland, NHS Health Research Authority. Does my project require review by a Research Ethics Committee? (2020) https://web.archive.org/web/20230110232131/https://www.hra-decisiontools.org.uk/ethics/docs/Algorithm%20-%20Does%20my%20project%20require%20REC%20review%20v2.0%2020200304.pdf [Accessed March 6, 2023]

154. NHS Health Research Authority. Standard operating procedures for Research Ethics Committees (version 7.4). (2019) https://www.hra.nhs.uk/about-us/committees-and-services/res-and-recs/research-ethics-committee-standard-operating-procedures/

155. NHS Health Research Authority. Research tissue banks and research databases. (2022) https://web.archive.org/web/20221204084259/https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/research-tissue-banks-and-research-databases/ [Accessed March 6, 2023]

156. NHS Health Research Authority. CAG pre-application checklist: Do I need to process identifiable information without consent? (2018) https://web.archive.org/web/20230110232148/https://s3.eu-west-2.amazonaws.com/www.hra.nhs.uk/media/documents/CAG_pre-application_checklist.pdf [Accessed March 6, 2023]

157. NHS Health Research Authority. Do I need to apply to the Confidentiality Advisory Group (CAG)? (2023) https://www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group/guidance-cag-applicants/

158. NHS Health Research Authority. NHS Health Research Authority. (2019) https://www.hra.nhs.uk [Accessed July 23, 2019]

159. NHS Health Research Authority. Research Ethics Service and Research Ethics Committees. (2023) https://www.hra.nhs.uk/about-us/committees-and-services/res-and-recs/

160. NHS Health Research Authority. Non-NHS Research Ethics Committees. (2018) https://www.hra.nhs.uk/about-us/committees-and-services/res-and-recs/non-nhs-research-ethics-committees/ [Accessed March 6, 2023]

161. NHS Digital. Hospital Episode Statistics. (2019) https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics [Accessed July 23, 2019]

162. NHS England. Access to patient records through the NHS App. (2023) https://digital.nhs.uk/services/nhs-app/nhs-app-guidance-for-gp-practices/guidance-on-nhs-app-features/accelerating-patient-access-to-their-record [Accessed March 6, 2023]

163. UK. Access to Health Records Act 1990. (1990) https://www.legislation.gov.uk/ukpga/1990/23/contents [Accessed July 6, 2021]

164. NHS Digital. Accelerating patient access to their record: Information for secondary, community and mental health service staff. (2022) https://web.archive.org/web/20221101191344/https://digital.nhs.uk/services/nhs-app/nhs-app-guidance-for-gp-practices/guidance-on-nhs-app-features/accelerating-patient-access-to-their-record/information-for-health-and-care-providers-working-in-secondary-care-community-care-and-mental-health-services

165. NHS Health Research Authority. Research summaries. (2023) https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/?query=19%2FES%2F0144&research_type=&rec_opinion=&date_from=&date_to=&relevance=true [Accessed March 6, 2023]

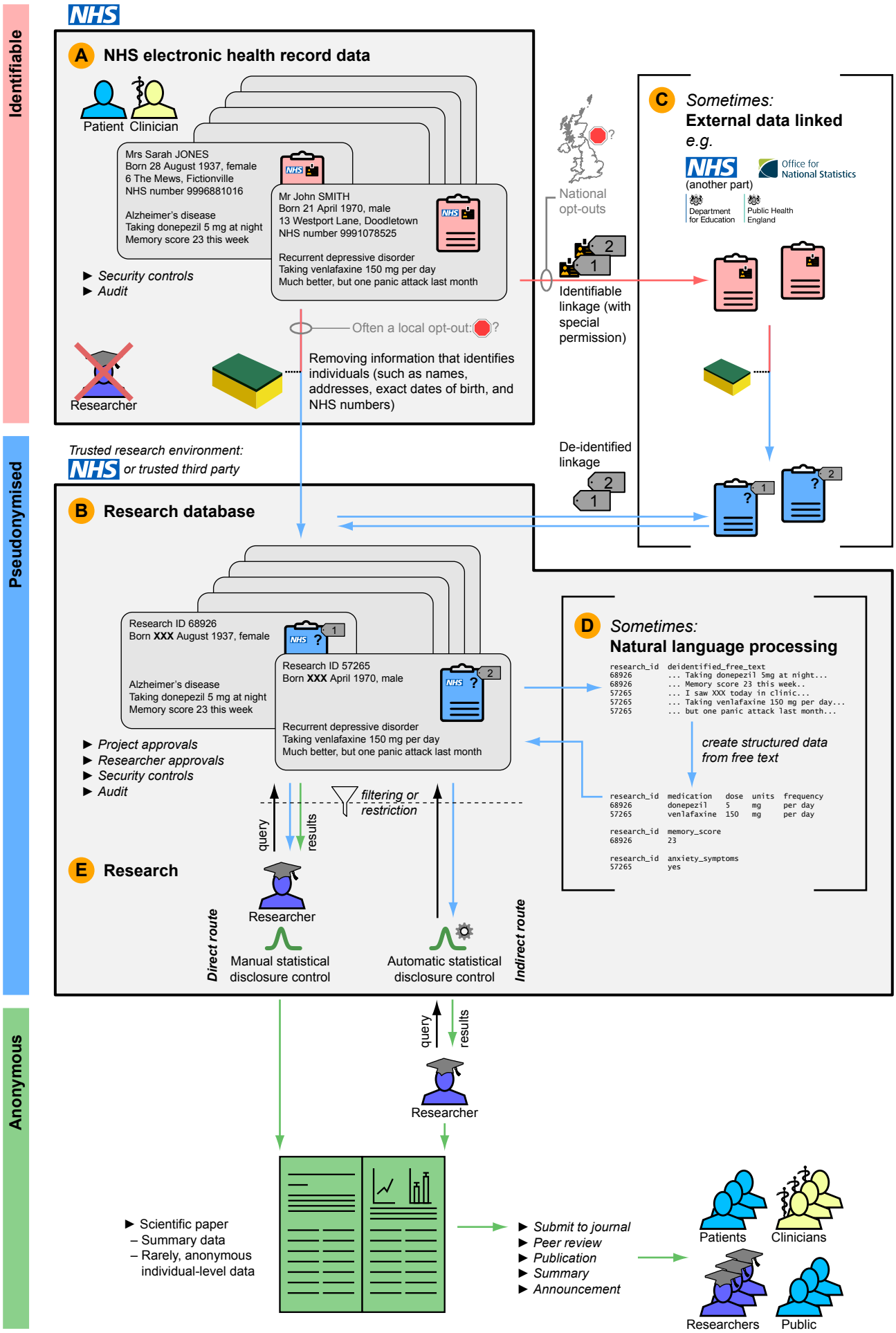
166. NHS Health Research Authority. Confidentiality Advisory Group registers. (2023) https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/confidentiality-advisory-group-registers/ [Accessed March 6, 2023]

167. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke J-P, Ford DV, Brown G, Leake K. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* (2009) 9:3. doi: 10.1186/1472-6947-9-3

168. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, Minion J, Boyd AW, Newby CJ, Nuotio M-L, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. *Int J Epidemiol* (2014) 43:1929–1944. doi: 10.1093/ije/dyu188

11 FIGURE LEGENDS

Figure 1: Typical journey of data from NHS electronic health records to research. Data colour coding follows **Supplementary Figure 1**. All examples are fictional and the fictional NHS numbers are from the official test range. **(A)** Typical starting point: identifiable health data held by the NHS (**Supplementary Table 1**). **(B)** De-identification creates a pseudonymised research database within a trusted research environment (TRE) (**Supplementary Table 2**), which might be within the NHS or provided by a trusted service (55,64,66,167). Local opt-outs may apply. **(C)** For some studies, additional data may be linked in, from external sources. There are identifiable and de-identified ways to do linkage (**Supplementary Table 2**). National opt-outs may apply. **(D)** Sometimes, de-identified free text is passed through natural language processing (NLP) software to extract more structured data (**Supplementary Table 2**). **(E)** Research can then take place, by approved researchers doing approved projects (studies). The researchers might be permitted to work within the TRE, either (a) by making electronic queries of the pseudonymised data (filtered/restricted by those queries and sometimes by other restrictions such as access only to partial data) and ensuring statistical disclosure control is performed (direct route, left), or (b) by using software to isolate them from the pseudonymised data itself (indirect route, right), e.g. (168). The latter is preferable (33). Either way, anonymous results are produced (**Supplementary Table 2**). The researchers write up their findings and submit them for peer review (**Supplementary Table 1**), and then the results are published and disseminated.


Identifiable
Pseudonymised
Anonymous
NHS
A NHS electronic health record data
Patient
Clinician
Mrs Sarah JONES
Born 28 August 1937, female
6 The Mews, Fictionville
NHS number 9996881016
Alzheimer's disease
Taking donepezil 5 mg at night
Memory score 23 this week
Mr John SMITH
Born 21 April 1970, male
13 Westport Lane, Doodletown
NHS number 9991078525
Recurrent depressive disorder
Taking venlafaxine 150 mg per day
Much better, but one panic attack last month
Security controls
Audit
Often a local opt-out:
Removing information that identifies individuals (such as names, addresses, exact dates of birth, and NHS numbers)
Researcher
National opt-outs
Identifiable linkage (with special permission)
C Sometimes:
External data linked
e.g.
NHS (another part)
Office for National Statistics
Department for Education
Public Health England
De-identified linkage
Trusted research environment:
NHS or trusted third party
B Research database
Research ID 68926
Born XXX August 1937, female
Alzheimer's disease
Taking donepezil 5 mg at night
Memory score 23 this week
Research ID 57265
Born XXX April 1970, male
Recurrent depressive disorder
Taking venlafaxine 150 mg per day
Much better, but one panic attack last month
Project approvals
Researcher approvals
Security controls
Audit
D Sometimes:
Natural language processing
research_id deidentified_free_text
68926 ... Taking donepezil 5mg at night...
68926 ... Memory score 23 this week..
57265 ... I saw XXX today in clinic...
57265 ... Taking venlafaxine 150 mg per day...
57265 ... but one panic attack last month...
create structured data from free text
research_id medication dose units frequency
68926 donepezil 5 mg per day
57265 venlafaxine 150 mg per day
research_id memory_score
68926 23
research_id anxiety_symptoms
57265 yes
filtering or restriction
query
results
E Research
Researcher
Direct route
Manual statistical disclosure control
Automatic statistical disclosure control
Indirect route
query
results
Researcher
Scientific paper
– Summary data
– Rarely, anonymous individual-level data
Submit to journal
Peer review
Publication
Summary
Announcement
Patients
Clinicians
Researchers
Public