

UNIVERSITY OF BIRMINGHAM

University of Birmingham
Research at Birmingham

Guidance for researchers wanting to link NHS data using non-consent approaches

Cross, Lauren; Carson, Lauren Emma; Jewell, Amelia; Heslin, Margaret; Osborn, David; Downs, Johnny; Stewart, Robert

DOI:

[10.23889/ijpds.v5i1.1355](https://doi.org/10.23889/ijpds.v5i1.1355)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Cross, L., Carson, L. E., Jewell, A., Heslin, M., Osborn, D., Downs, J & Stewart, R 2020, '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', *International Journal of Population Data Science*, vol. 5, no. 1, 34. <https://doi.org/10.23889/ijpds.v5i1.1355>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Cross, L., Carson, L. E., Jewell, A., Heslin, M., Osborn, D., Downs, J. and Stewart, R. (2020) "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: A thematic analysis of feedback from the Health Research Authority Confidentiality Advisory Group", *International Journal of Population Data Science*, 5(1). doi: 10.23889/ijpds.v5i1.1355.

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

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

Lauren Cross^{1*}, Lauren Emma Carson¹, Amelia Jewell², Margaret Heslin³, David Osborn⁴, Johnny Downs¹, and Robert Stewart¹

Submission History

Submitted:	08/06/2020
Accepted:	30/07/2020
Published:	02/10/2020

¹Department of Psychological Medicine, King's College London, Strand, London WC2R 2LS, UK

²South London and Maudsley NHS Foundation Trust, National Institute for Health Research (NIHR), Maudsley Biomedical Research Centre, Denmark Hill, London, SE5 8AF, UK

³Department of Health Services & Population Research, King's College London, Strand, London WC2R 2LS, UK

⁴Division of Psychiatry, University College London, Maple House, 149 Tottenham Court Rd, Bloomsbury, London W1T 7BN, UK

Abstract

Introduction

The use of linked data and non-consent methodologies is a rapidly growing area of health research due to the increasing detail, availability and scope of routinely collected electronic health records data. However, gaining the necessary legal and governance approvals to undertake data linkage is a complex process in England.

Objectives

We reflect on our own experience of establishing lawful basis for data linkage through Section 251 approval, with the intention to build a knowledgebase of practical advice for future applicants.

Methods

Thematic analysis was conducted on a corpus of Section 251 feedback reports from the NHS Health Research Authority Confidentiality Advisory Group.

Results

Four themes emerged from the feedback. These were: (a) Patient and Public Involvement, (b) Establishing Rationale, (c) Data maintenance and contingency, and the need to gain (d) Further Permissions from external authorities prior to full approval.

Conclusions

Securing Section 251 approval poses ethical, practical and governance challenges. However, through a comprehensive, planned approach Section 251 approval is possible, enabling researchers to unlock the potential of linked data for the purposes of health research.

Keywords

data linkage; section 251; thematic analysis; non-consent approaches

Highlights

- Gaining Section 251 support can be a lengthy and complex process and there is partial guidance to help researchers navigate the ethical, practical and governance issues needed to gain approval.
- This study aims to synthesise available information using a corpus of existing S251 feedback from previous applications, and to build a knowledge base that will support future applicants.

- Four themes emerged from the feedback. These were: (a) Patient and Public Involvement, (b) Establishing Rationale, (c) Data maintenance and contingency, and (d) Further Permissions from external authorities prior to full approval.

Introduction

Informed consent is a fundamental principle governing the use of patient identifiable information within health research [1, 2].

*Corresponding Author:

Email Address: lec67@medschl.cam.ac.uk (Lauren Cross)

The research was conducted whilst Lauren Cross was employed at King's College London.

It is recognised, however, that there are situations where gaining informed consent may undermine the ability to conduct research in the public interest. Consent requirements can cause systematic response biases leading to unrepresentative samples which challenge the validity of research findings, and possible over or underestimation of risk-outcome effects [3]. In addition, large scale participant recruitment and measurement of research variables can simply prove too costly for participants and research funders to sustain using conventional cohort methodologies [4].

In England and Wales, Section 251 (S251) of the NHS Act 2006 provides statutory power to allow NHS patient identifiable information needed to support essential research activity to be used without individuals' consent. Following approval, patient identifiable information can be transferred without breach of common law duty of confidentiality, enabling novel large-scale research which might not be appropriate or achievable using traditional consented approaches. Recent examples of using linked psychiatric electronic health records linking with external data sources include health services research studies examining acute hospital admission patterns in patients with severe mental illness [5], falls and fractures in people with severe mental illness [6], and school absence patterns in young people receiving specialist mental health care [7].

Although S251 can facilitate large data resources, and highly powered large-scale research studies, it is essential that applications are approached carefully and responsibly. Securing the correct approvals for data linkage can be a lengthy and complex process in England [8]. Application guidance is limited, and although some application guidance is provided it is not always in a user-friendly format and genuine insight is often gained through a 'trial-and-error' basis. The aim of this study was to synthesise available information using a corpus of existing S251 feedback from previous applications for approval, and to build a knowledge base that will support future applicants. It should also be noted that the common shorthand "S251 approval" is not an accurate representation of the process discussed in this analysis. Approval is given under the Health Service (Control of Patient Information) Regulations 2002, and S251 NHS Act 2006 creates the authority for these Regulations.

This review has been conducted as part of King's College London's (KCL) Medical Research Council Pathfinder programme (MC_PC_17214) to replicate and facilitate mental health electronic records data linkages with the intention of creating a national network of linked data. To our knowledge, it is the first study of its kind.

Methods

Information were gathered from a corpus of nine documents related to research applications to link data for research including: four Confidentiality Advice Team (CAT) feedback forms and five Confidentiality Advisory Group (CAG) outcome letters from S251 applications. The CAG offer independent and expert advice on S251 approval to the Health Research Authority and Secretary of State for Health, whereas the CAT provides pre-application assessment prior to formal scrutiny by

the CAG. Not all study applications are reviewed by the CAT team. Therefore, one application had an outcome letter but did not have a CAT feedback form.

All research applications were made by, or in collaboration with*, the South London and Maudsley NHS Foundation Trust (SLaM) issued between May 2017 and July 2019. Two applications (Documents 1–4) were made prior to, and three applications after The Data Protection Act 2018.

SLaM is a large mental health Trust covering a geographic catchment of four London boroughs (Lambeth, Southwark, Lewisham, and Croydon) and a diverse population of around 1.2 million residents. It is the main provider of comprehensive adult mental health care to its catchment area population, as well as providing prison in-reach and a number of national specialist services. SLaM has an extensive track record in data linkage, developing means and expertise to derive research output from linked electronic mental health records through the Maudsley Biomedical Research Centre (established with NIHR funding in 2007) and longstanding partnerships with King's College London.

The S251 research applications included in this review sought approval to link electronic mental health records data to a variety of existing health and non-health administrative data resources including: Department for Work & Pensions employment and benefits data, maternity and neonatal/paediatric electronic health records data, Children and Family Court Advisory and Support Service data around care proceedings, Hospital Episode Statistics data concerning physical health, and HIV and AIDS electronic health record data. All Trusts and bodies have been anonymised within the text to retain anonymity.

Data was analysed using thematic analysis (Braun, 2012) and coded using NVivo software. The thematic analysis involved 5 key steps: i) familiarisation with the data, through repeated readings of the transcripts during which thoughts and potential codes were noted; ii) generation of initial codes, through intense open-coding of data to generate an initial coding frame based on thematic categories rooted in the data; iii) identification of themes, through a detailed review of the coding frame to sort codes in to potential themes; iv) review of themes, through refinement of the developing themes; v) definition and refinement of themes, through detailed exploration of relationships within and between codes, merging and pruning of codes, and revision of thematic definitions [9]. Applications were initially coded by one researcher (LC). Findings and emerging themes were tested for validity through one to one and group discussions with the interdisciplinary team of co-authors, who have backgrounds in data science, clinical epidemiology and NHS information governance. This was done through team discussion rather than independent cross-referencing as the documentation had previously been reviewed by all team members. Further, this analysis is unusual as documentation, rather than participant interview transcripts, were analysed.

Results

Of the five applications submitted, four were provisionally supported and one outcome deferred (although supported

*One application was submitted by SLaM on behalf of another NHS mental health trust.

in principle). Four themes emerged from the feedback given by the CAT and CAG. These were: (a) Patient and Public Involvement, (b) Establishing Rationale, (c) Data maintenance and contingency, (d) Further permissions (See Table 1). In addition, ideas concerning clarity, consistency and detail were evident across all themes within the text.

Patient and Public Involvement (PPI)

The importance of engagement with patients was one of the most predominant themes within the feedback, and explicitly stated to be “*considered an important factor*” in more recent outcome letters (e.g. Document 4, P4, L39). Within this theme four sub-themes were present, highlighting that patient engagement work should be: embedded, evidenced, targeted and accessible.

Embedded

Analysis of the text highlights that PPI work should be present throughout the research cycle “*members acknowledged that this group had been consulted in the design phase of the project and would continue to be engaged with as the project progressed, which was commended*” (Document 5, P4, L20). The emphasis here was on an approach to PPI that was beyond tokenistic, forming an integral part of the research schedule and strategy. In particular, there appeared to be two phases of the research cycle in which PPI is viewed as of importance.

First, during the initial stages (prior to linkage) the CAG expected methodological consideration from service users “*feedback from the planned activity would also need to be reported to understand the views of this cohort in relation to the proposal*” (Document 3, P5, L30); and specific focus on the use of patient identifiable information during the linkage process “*further information would be required in this area to confirm that the service user group understood that the proposed linkage would involve the disclosure of confidential patient information to another organisation*” (Document 5, P5, L9).

Second, the CAG required the development of information materials to raise the profile of a potential linked database highlighting its use for research activities once the intended

linkage is complete. This includes informing patients and the public about how their data is being used, and importantly the opportunity and tools to object (opt-out) “*The Group was satisfied that the patient notifications and dissent mechanisms offered were appropriate and adequate for the project*” (Document 5, P4, L40). Such information needs to be released with enough time ahead of the linkage to allow a “*specific time period for meaningful opt-out*” (Document 1, P3, L27).

The CAG strongly endorsed channels and mechanisms for dissent or opt-out to remain up-to-date and clearly communicated to the public. Applications without clear opt-out / dissent mechanisms were asked to be amended “*It was commented that the posters should be revised to include information around patient opt-out*” (Document 4, P5, L11). Such communication can, however, be successfully achieved through a number of different channels including: public-facing websites, posters and leaflet materials containing clear guidance regarding notification and dissent.

All of the applications relied on local opt-out mechanisms, however as the NHS national opt-out scheme (whereby, in the UK, any patient can opt out of their NHS data being used for secondary purposes now or in the future) becomes mandatory [10], future applications will have to consider both national and local governance systems to enable sufficient notification and dissent.

Evidenced

It was clear from the text that evidence of PPI work needs to be explicit and contained within the appendices of an application. Where detail was not provided, or documents were missing, this was requested “*this poster had not been included in the submission. CAG asked that it was provided*” (Document1, P4, L40), although draft forms and intended updates were deemed acceptable evidence. This means that communication materials do not have to be already approved and in circulation prior to application.

As previously highlighted, in the context of service user groups, the CAG were interested in clear and explicit confirmation of the acceptability of use of patient identifiable data during the linkage process. We illustrate this, below from the PPI minutes which were submitted to, and

Table 1: Summary table of themes and subthemes

Themes
1. Patients and Public Involvement (PPI) Subthemes: a) Embedded b) Evidenced c) Targeted d) Accessible
2. Establishing rationale Subthemes: a) Establishing appropriate grounds b) Establishing appropriate and proportionate use of data
3. Data maintenance and contingency
4. Further permissions



subsequently commented on by the CAG within the feedback: “[The researcher] described the process of the linkage and emphasised that although the linkage uses identifiers like name and date of birth, no health data leaves SLaM and health data is only available to researchers once identifiers have been removed and the linkage is complete” (Document 10, P2, L10). “Group feedback: Linkage process – No comments. Everyone happy with proposed process” (Document 10, P2, L29). Therefore, it is important that applicants approaching PPI work with service users, incorporate explicit questioning and careful minute-taking of this to support their S251 application.

Targeted

Feedback from the CAT and CAG highlighted that, in order for PPI work to be meaningful, it needed to be project-specific and discussed with those who were likely data subjects “it is unclear from the detail provided whether there has been any specific engagement involving patients with [the specific disorder]” (Document 6, P2, L20). In addition, it requires targeted communication pathways “The project was relying on the established communications strategy for the [research database] within [NHS Trust 1] as the project notification and dissent mechanism. Members agreed that, due to the sensitivities around the information which would be linked for the project, specific information should be displayed on the CRIS website, to enable a project-specific objection mechanism to be operated” (Document 3, P5, L35).

Accessible

Analysis revealed that a sensitive approach to PPI is required, with information accessible to all users. First, it is important that any information is easily found and acted upon by patients and the public “The Group was of the opinion that the patient’s right to opt-out was ... difficult to locate when reviewing the information online” (Document 2, P5, L27). This is inclusive of control groups where the intention is to use individual-level data “The Group agreed that the control group needed to be informed about the study and given the opportunity to dissent, unless the data was available only in aggregated form” (Document 1, P5, L3). Accessing control groups from the general population who do not use mental health services is acknowledged as challenging; however, efforts should be made to explore existing engagement pathways already in operation for the administrative database intended to be linked.

Moreover, a variety of communication mechanisms were recommended in order to maximise information promotion and response pathways “when offering dissenting options to patients, it was preferred that a number of communication modes are provided, i.e. telephone, email and postal” (Document 4, P5, L11). Therefore, applicants should aim to develop a range of different mechanisms and formats to deliver information to patients and the public. Notably, social media was only briefly mentioned within one document (Document 9, P2, L18); therefore, given the increasing use of digital tools within the public domain [11], this is a potentially under-used (or at least under-acknowledged) resource within PPI work for S251 approval. However, it is acknowledged that the use

of digital tools for PPI work has its own biases as certain community groups may have limited access to these platforms to provide their feedback.

Finally, clarity of information was highlighted as a central characteristic of successful PPI work. Patients and the public should not only have access to information, but it should be written in a digestible format for a non-expert reader “members also commented that the information should clearly explain the governance arrangements for the project, to ensure that it was clear that identifiable information would only be used to facilitate linkage between the data sources and advise that analysis would be undertaken on an anonymised dataset” (Document 3, P6, L4). This is essential as patients need knowledge to make an informed opinion as to whether they wish to dissent or disagree with the linkage. Applicants should aim to include plain English summaries of all technical aspects of the data linkage and access arrangements to interpret and explain text as appropriate. This will promote understanding and is essential when presenting to service user groups, and designing promotional material. However, there are other challenges for undertaking meaningful PPI including time constraints and financial implications [12]. These issues can potentially be overcome through the use of pre-existing PPI networks.

Establishing rationale

The second most predominant theme within the documents and feedback from the CAG was the need to justify research and methodological rationale for projects. Given the potential sensitivity and nature of S251 as a mechanism for allowing data use without individual consent, it was not surprising that identifying need and establishing a clear project framework was a key concern for those granting approval. This was captured within the text through the themes of justification of appropriate: grounds, use of data, and security protocol.

Establishing appropriate grounds

One of the more important aspects of securing authorisation from the CAG was establishing a lawful basis for S251 approval. This means that applicants should be able to justify that a) the purpose of the project is for medical research and in the public interest, and b) the proposed methodology is appropriate and proportionate to the study objectives. These principles are aligned to the Data Protection Act (DPA) 1998 & 2018.

This was most effectively achieved in applications through explicit mention that certain GDPR exception conditions have been met. For example, “In terms of Schedule 3, we consider that Condition 8 has been met. (1) The processing is necessary for medical purposes and is undertaken by- (a) a health professional, or (b) a person who in the circumstances owes a duty of confidentiality which is equivalent to that which would arise if that person were a health professional. (2) In this paragraph “medical purposes” includes the purposes of preventative medicine, medical diagnosis, medical research, the provision of care and treatment and the management of healthcare services” (Document 8, P2, L34), accompanied by sound academic justification “members noted that the applicants have provided a number of sound justifications

which supported the requirement for research in order to gain a better understanding of the mental health needs of mothers" (Document 3, P3, L13). Applicants are thus expected to pair a research rationale with meeting specific conditions of the DPA in order to clearly identify that the proposal is appropriate for medical purpose.

Regarding justification for non-consent methodologies, sample size and bias were the most referenced arguments. In the context of sample, the CAG were receptive to the notion that in cases where the target population was of a considerable size it would not be practical nor possible to capture the entire population through traditional methodologies. Furthermore, there was acknowledgement that attempts to recruit participants de novo could incur biases leading to groups who are harder to reach being systematically excluded from the analyses. This is particularly relevant to mental health research, where those who have greater need and complex clinical and social comorbidities are less likely to be represented in traditional research.

Establishing appropriate and proportionate use of data

The CAG requested clear and concise detail on exactly which data were being requested. This encompassed explicit detail on the scope: "Will a retrospective cohort be included within the database (i.e. back to 2005 at NHS Trust 2 and 2014 at NHS Trust 3 when BadgerNet was introduced)? If so, provide further details of this cohort" (Document 8, P1, L17); and the size of the cohort "Clarify the anticipated size of the patient cohort to be included in the database (mothers and babies)" (Document 8, P1, L20). This attention to detail enables the CAG to establish exactly what the linkage will look like, and the information available to researchers. However, the rationale behind why sensitive information is necessary is also important "Part B Q4 – states that date of death is required for analysis- provide clarification around why this is required. . ." (Document 9, P2, L12). Therefore, those applying for S251 approval should be able to justify not only precisely which data they wish to be linked, but how it will be used. This reflects a clear concern that data are to be used appropriately: "confirm that patients' country of birth will be used for linkage purposes only and not included in the analysis of the data" (Document 1, P6, L16).

Justifying appropriate security protocol

Security of procedures and personnel was an essential component of securing S251 approval. For example, data transfer along a secure pipeline was also cited as important: "Clarify how and to who data will be transferred" (Document 9, P1, L20). This should be explicit "patient identifiers will be sent via an N3 connection" (Document 9, P1, L29); and adhere to a secure transfer protocol (a system in which data are accessed, managed and transferred via an assured data pipeline).

Furthermore, all sites involved in the data linkage process (whether processing or hosting) require adequate demonstration of security adherence. It was a requirement of all applicants to achieve a grading of 'standards met' on the NHS Data Security and Protection (previously Information Governance) Toolkit submission (NHS Digital, 2019) to evidence internal

data storage and access procedures, and to promote legitimate and safe use of data. For example, "Specific Conditions of Support (Provisional). . . 2. Confirmation provided from the IG Delivery Team at Health Organisation 1 to the CAG that the relevant Data Security and Protection Toolkit (DSPT) submission(s) has achieved the 'Standards Met' threshold. See section below titled 'security assurance requirements' for further information." (Document 6, P6, L24).

The CAG also emphasised the importance of ensuring all personnel handling data had appropriate clearance, typically contractual agreements were used to evidence compliance "This honorary contract ensures that all individuals working on the data have a duty of confidentiality" (Document 8, P2, L34). This provides assurance that all personnel uphold security standards and the duty of confidentiality and care.

In particular, as they have temporary access to Patient Identifiable Information (PII), there were concerns regarding the specifics of personnel conducting the matching "It states that NHS Trust 1 identifiers will be transferred to a named member of staff at the non-health governmental department who will conduct the matching of identifiers; however, later in the form (Part B, Q3) it states that identifiers will be accessed by a small informatics team at the DWP. Clarify how and to who data will be transferred and who will have access to this identifiable information held at DWP for linkage" (Document 9, P1, L15). Applicants for S251 approval should therefore have a clear and evidenced policy and approval procedure to ensure that all staff handling data (either at point of linkage or analysis) do so appropriately and with care. Furthermore, due to the sensitive nature of data linkage, keeping the number of individuals involved in handling matching proceedings to a minimum, and being able to explicitly name such individuals, will strengthen applications. This acts to minimise the scope for data misuse and ensures accountability for any security breaches.

Data maintenance and contingency

A small but noticeable theme from the feedback text highlights the importance of legacy planning for the future of the database once linkage is complete. This was cited as needing to include fiscal and human resources to support data management "it was unclear what would happen to the study data if the additional funding was not received" (Document 4, P4, L17); to identify a clear time-line "Clarify how long the request for support is anticipated to last" (Document 9, P1, L11); and to clarify an exit strategy, as appropriate: "Can you confirm the intended exit strategy from support under the Regulations- i.e. will Health organisation 1 be destroying the file including confidential patient information, and the timeframe for this" (Document 7, P3, L15). Therefore, it is important that researchers carefully consider the maintenance and legacy of the database after its creation. Having a clear picture of what will happen next, and explicitly when confidential data will be destroyed (and by whom) is essential.

Further permissions

This is an acknowledgement that S251 is only one part of the process and additional permissions are required to signify approval prior to sign-off. In the UK, as stated in Regulation 5 of the Health Service (Control of Patient

Information) Regulations 2002, approval of an NHS Research Ethics Committee (REC) is a requirement for confidential patient information processed under the Regulations for medical research. Therefore, NHS REC approval is an essential requirement to permit research involving NHS patients, and those in social care, and ensure the research is independently reviewed against an ethical framework. Applicants should note that S251 can be granted *pending* REC approval, therefore it is possible to secure provisional S251 approval before REC approval is in place.

In addition, Caldicott approval was included within supporting documentation of all applications in the form of a signed letter. Caldicott Approval refers to permissions granted by the Caldicott Guardian. Caldicott Guardians are typically a senior member of staff within a health or social care organisation, tasked with ensuring personal information about those who use its services is used legally, ethically and appropriately in order to maintain confidentiality [13]. However, it only provides governance assurance for health data. There are no Caldicott equivalents for administrative non-health data (for example a linkage with the Department for Education's National Pupil Database, which comprises school's data), and whilst additional approval is sought from non-health authorities by the CAG there doesn't appear to be a consistent gold standard as to which form approval should take.

Discussion

The aim of this study was to summarise feedback from the NHS Health Research Authority Confidentiality Advisory Group on S251 applications in England. To our knowledge, this is the first exploration of its kind, providing valuable insights to support future applicants wanting to use routinely collected, linked NHS data for the purpose of research.

From the findings we derive the following recommendations for future S251 applications:

1. Applicants should undertake substantial PPI work prior to application, and continue to engage with patients and the public throughout the research cycle. This should be comprehensive, project-focused and broadly accessible to both service users and control groups. Particular efforts should be made to engage with those whose data will be used, and to communicate opt-out procedures.
2. Careful and clear documentation of PPI work should be included in appendices to support the application. As well as views on the rationale for the proposed linkage and its intended use, applicants should carefully evidence explicit confirmation of patients giving methodological approval for the use of patient identifiable information within the linkage process as well as notification and dissent mechanisms within data-base promotion materials.
3. Applicants should build a strong case, rooted in evidence, that S251 approval is appropriate and indeed necessary. Where GDPR conditions have been met, this should be explicitly stated to establish a legal basis for deviating from common law duty of confidentiality.

4. Applicants should carefully consider security when putting together proposals. This should be inclusive of both technical and human factors. As such, applicants should seek to go beyond following the 'letter of the law' but embed protocols to exceed this- facilitating robust security procedures to protect patient identifiable information and sensitive data.
5. Applicants should establish a legacy for the maintenance of the linked database in terms of finances and data management, and an exit strategy for the secure destruction of identifiable data within a suitable timeframe.
6. Applicants should not rely on the S251 in isolation in order to satisfy legal requirements to link data. Approval should also be sought from additional authorities (including gaining REC and Caldicott assurance) to permit appropriate and safe use of health data and associated linkages to other data sources.
7. Finally, careful attention to detail is necessary for a successful S251 application.

As far as we are aware, there is no academic literature on this topic. This makes it difficult to verify findings, although our experiences and recommendations are similar to those given by the CAG [14]. However, our analysis is novel in providing a practical framework to assist in interpreting guidance and supplementing advice provided by the CAG. As identified by our analysis, the CAG advice highlights the need for proportionate PPI. This includes engaging patients and public where feasible and providing clear object and dissent mechanisms. However, the CAG does not provide specific details on how to achieve this in practice. Similarly, establishing rationale is also a prominent theme identified within both our analysis and advice provided by the CAG. This highlights that applicants will need to satisfy the legal requirements established under S251; medical purposes, participant consent unable to be reasonably sought, the proposed research is in the public interest, and no reasonable alternative. As in the context of PPI, the CAG does not however provide detail on what may be considered reasonable or indeed appropriate and proportionate use of data beyond minimizing the identifiers requested for linkage. Furthermore, establishing security protocol is a theme present both within our analysis and guidance provided by the CAG. Whilst the CAG's advice largely focuses on human resource aspects of data security including contractual obligation detailed within our analysis, an interesting piece of practical advice highlighted within the CAG's guidance but omitted from our analysis, is the preference for a third party to undertaking the data linkage utilizing pseudonyms or a data-linkage key. However, there is no further detail provided regarding specific technical security standards deemed acceptable for accessing confidential patient information for the purpose of research within the CAG's standardized advice pages beyond validation through IG toolkits.

Our major themes, data maintenance and contingency, and further permissions were also consistent with CAG published advice. As in our analysis, the CAG outlines the need to consider the legacy of the data and putting in steps to withdraw support and highlight the need for additional

approval. Similar to our findings, REC approval is explicitly mentioned however Caldicott Approval is not. Therefore, Caldicott Approval may strengthen an application but is not necessary for permission to be granted by the CAG. Our findings were also consistent with CAG advice in the context of non-health data, however whilst establishing lawful basis and securing permission were stated as essential, neither our analysis nor the CAG's advice provide a gold standard for implementation practice. Therefore, establishing a non-health data alternative is potentially a necessary and useful area of future attention.

Limitations & future research

Findings and recommendations should be interpreted with caution, bearing in mind certain limitations with our approach. In particular, the analysis was based on a relatively small and select sample of applications for approval for data linkage produced under the auspices of a single site. Therefore, commentary and feedback may be specific to the experiences, strengths and weaknesses of this organisation. However, although the sample was select, there was heterogeneity across research topics, proposed linkages and methodologies. We are also aware that not all the themes generated by our analyses will be applicable to all future S251 applications and caution must be used in incorporating all our recommendations. Another limitation of this study is the use of thematic analysis to explore this topic. Although thematic analysis is very good for a broad overview of the data and can be used on documents in addition to research interviews, it is not nuanced enough as a methodological tool to explore the language used nor is it sophisticated enough to explore a particular phenomenon.

Furthermore, as far as we are aware, there is no academic literature in this area. This makes it difficult to verify findings. However, our experiences and recommendations do appear consistent with advice given by the CAG [14]. Any conclusions should therefore be considered as first steps in building a knowledge base regarding applying for health record linkage permissions, and as a stimulus for further exploration and collation of experience to help cross-validate and extend findings.

Conclusion

Gaining Section 251 support can be a lengthy and complex process. To date there has been limited guidance that helps researchers navigate the ethical, practical and governance issues needed to gain approval. The analysis and recommendations within this article provide reasoned, evidenced, and clear approaches to preparing future applications. We hope it supports both future applicants and the CAG as the recommending authority, to continue high quality research using existing NHS data resources and enhance the nation's health and social care.

Acknowledgments

We are grateful to our colleagues and all applicants for making their CAT and CAG feedback reports available, enabling

this present analysis. In addition, we thank Professor Ruth Gilbert, University College London, for her valued thoughts and feedback on this article.

Funding statement

LC and LEC were supported by a Medical Research Council Mental Health Data Pathfinder Award to King's College London. LEC also received salary support from the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. AJ and RS are part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust and King's College London. JD is supported by NIHR Clinician Science Fellowship award (CS-2018-18-ST2-014) and has received support from a Medical Research Council (MRC) Clinical Research Training Fellowship (MR/L017105/1) and Psychiatry Research Trust Peggy Pollak Research Fellowship in Developmental Psychiatry. MH reports funding by NIHR and the Stefanou Foundation, UK. DO is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at University College London Hospitals (UCLH). DO is also supported by the National Institute for Health Research ARC North Thames. This report is independent research supported by the National Institute for Health Research ARC North Thames. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

Ethics statement

Due to the secondary nature of this research, ethical approval was not necessary for this analysis.

Statement on conflicts of interest

No authors have any conflicts to declare.

References

1. The Caldicott Report. IHRIM. 1999;40(2):17–9.
2. Crook MA. The Caldicott report and patient confidentiality. *Journal of Clinical Pathology*. 2003;56(6):42610. [1136/jcp.56.6.426](https://doi.org/10.1136/jcp.56.6.426).
3. Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. *BMJ (Clinical research ed)*. 2009;338:b866–b10. [1136/bmj.b866](https://doi.org/10.1136/bmj.b866).
4. Wadman M. Child-study turmoil leaves bitter taste. *Nature*. 2012;485(7398):287–810. [1038/485287a](https://doi.org/10.1038/485287a).
5. Jayatilleke N, Hayes RD, Chang CK, Stewart R. Acute general hospital admissions in people

- with serious mental illness. *Psychological medicine*. 2018;48(16):2676–8310. [1017/s0033291718000284](https://doi.org/10.1017/s0033291718000284).
6. Stubbs B, Mueller C, Gaughran F, Lally J, Vancampfort D, Lamb SE, et al. Predictors of falls and fractures leading to hospitalization in people with schizophrenia spectrum disorder: A large representative cohort study. *Schizophr Res*. 2018;201:70–810. [1016/j.schres.2018.05.010](https://doi.org/10.1016/j.schres.2018.05.010).
 7. Downs JM. Big data approaches to investigating Child Mental Health disorder outcomes. 2018.
 8. Downs JM, Ford T, Stewart R, Epstein S, Shetty H, Little R, 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(1):e02435510. [1136/bmjopen-2018-024355](https://doi.org/10.1136/bmjopen-2018-024355).
 9. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77–10110. [1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa).
 10. NHS Digital. National data opt-out operational policy guidance document 2019 [Available from: <https://digital.nhs.uk/services/national-data-opt-out/operational-policy-guidance-document>].
 11. Office of National Statistics. Internet users, UK: 20192019. Available from: <https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2019>.
 12. Jewell A, Pritchard M, Barrett K, Green P, Markham S, McKenzie S, et al. The Maudsley Biomedical Research Centre (BRC) data linkage service user and carer advisory group: creating and sustaining a successful patient and public involvement group to guide research in a complex area. *Research Involvement and Engagement*. 2019;5(1):2010. [1186/s40900-019-0152-4](https://doi.org/10.1186/s40900-019-0152-4).
 13. Roch-Berry C. What is a Caldicott guardian? *Postgrad Med J*. 2003;79(935):516–810. [1136/pmj.79.935.516](https://doi.org/10.1136/pmj.79.935.516).
 14. Health Research Authority. Guidance for CAG applicants 2019 [Available from: <https://www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group/guidance-confidentiality-advisory-group-applicants/>].

Abbreviations

CAT:	Confidentiality Advisory Team
CAG:	Confidentiality Advisory Group
DPA:	Data Protection Act
GDPR:	General Data Protection Regulations
PII:	Patient Identifiable Information
PPI:	Patient and Public Involvement
REC:	Research Ethics Committee
S251:	Section 251

