



BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Data sharing and publications from research databases given a favourable opinion by UK research ethics committees

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-039756
Article Type:	Original research
Date Submitted by the Author:	24-Apr-2020
Complete List of Authors:	Trace, Samantha; University of Portsmouth, School of Healthcare Professions Bracher, Mike; University of Southampton, Health Sciences Kolstoe, Simon; University of Portsmouth, School of Healthcare Professions
Keywords:	QUALITATIVE RESEARCH, ETHICS (see Medical Ethics), AUDIT, Information management < BIOTECHNOLOGY & BIOINFORMATICS, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Supérieur (ABES).

Data sharing and publications from research databases given a favourable opinion by UK research ethics committees

Samantha Trace¹, Mike Bracher² and Simon Kolstoe^{1,3}

¹ School of Health Care Professions, University of Portsmouth, PO1 2UP, UK

² School of Health Sciences, University of Southampton, SO17 1EN, UK

³ Correspondence to: Dr Simon Kolstoe; simon.kolstoe@port.ac.uk

Abstract

Objective To determine the level of data sharing and number of publications coming from research databases that have been given a favourable opinion by UK NHS research ethics committees.

Design Cohort study

Inclusion Criteria & Setting All research databases listed on the UK Health Research Authority's Assessment Review Portal (HARP) that had received a favourable ethics opinion as of January 2018.

Main Outcome Measures Publications and data access requests either listed on HARP or notified through subsequent email correspondence.

Results Out of 354 eligible databases, 34% had granted access requests and 40% had produced at least one peer reviewed paper or conference abstract/talk. We could not establish contact with 9% of databases, and 19% reported no access requests or publications. Only 9% of databases were up to date with all annual reports. Email responses from database owners showed a range of attitudes towards data sharing.

Conclusion Less than half of Research Databases that have received a favourable opinion from NHS research ethics committees share their data and produce publications. There is also considerable variability in the operation of Research Databases and understanding of the purpose of Research Databases. This work was hampered by incomplete records due to researchers not submitting annual reports.

Key words

Research Database; Research Ethics; Data Sharing; Publication;

Word Count:

3,822

Article Summary

Strengths and limitations of this study

- By using the UK Health Research Authority’s “HARP” database we were able to identify all research databases using National Health Service (NHS) data that were registered as of January 2018
- We were able to identify both publications and access requests from the majority of databases
- We identified numerous incomplete records in the HARP database
- Research teams were not consistent in their definition of a research database, and it is likely that many relevant databases may not be registered with the HRA

Introduction

As data analysis processes continue to evolve, research databases represent increasingly important resources within healthcare research, yet there is evidence that they are currently under-utilised[1]. In the UK a research database is defined as:

“...a structured collection of individual-level personal information, which is stored for potential research purposes beyond the life of a specific research project with defined endpoints. Research purposes in this context refers to analysis of data to answer research questions in multiple projects.”[2]

The Health Research Authority (HRA) is the administrative body that convenes and organises research ethics committees (RECs) authorised to review studies involving human participants that take place within the National Health Service (NHS), as well as falling under certain legislation[3]. Although most of the HRA’s functions applies to research undertaken in England, its role coordinating policy, and managing the Integrated Research Application System (IRAS), gives it close links to the other devolved UK nations (Scotland, Wales and Northern Ireland) including access to records for audit and service improvement purposes. Through IRAS, the HRA flags research database applications, and provides a specific question set for researchers wishing to have their arrangements for collection, storage and use of data reviewed (including arrangements for release of non-identifiable data for analysis by external researchers). This requirement is outlined in the UK wide GAfREC (Governance Arrangements for Research Ethics Committees) policy whereby the:

“...collection of personal information from past or present users of health or social care services, or use of previously collected information from which individual users of these services could be identified, either directly from that information or from its combination with other information in, or likely to come into, the possession of someone to whom the information is made available”

always requires an ethics review, however the review of more generalised database projects by ethics committees:

... may have benefits by facilitating programmes of research using information on human subjects without a need for specific project-based applications. Applicants

may seek generic ethical approval extending to specific projects undertaken using the data, subject to conditions agreed with the REC.”[4]

Consequently, in the UK, research databases differ from other types of research projects in that they are normally intended to be used multiple times, over a longer period of time, and perhaps by different research teams wishing to test a variety of hypotheses. When reviewing research databases RECs therefore consider the access arrangements being made available to researchers wishing to interrogate the database, including arrangements for subsequent publication of research results. Indeed, there is an implicit assumption that research databases will be used to generate many more publications than a normal research project.

In order to test this assumption, and to benchmark UK performance with other national studies[1], the HRA invited us to audit UK research database applications made through the IRAS system. This request formed part of the wider “Transparency Agenda” being pursued as a statutory duty by the HRA, but further encouraged by organisations such as the AllTrials campaign[5] and the REWARD Alliance[6]. A previous audit by the HRA showed that only one third of regular projects reviewed by RECs publish their results[7], raising a subsequent concern that research database projects may also be underperforming in terms of publishing outputs.

Methods

Research Databases were identified by searching the HRA Assessment Review Portal (HARP) using the initial inclusion criteria of research database applications on the system as of 1st January 2018. The number of eligible databases were then reduced using the following criteria:

- Favourable Ethics opinion
- Not a duplicate record or renewal request
- Not a Welsh application

A Microsoft Access database was created with an entry for each research database and information contained on HARP along with any uploaded annual or final reports were used to populate the database fields listed in Table 1 of the supplementary information. Following creation of the Access Database, primary contacts for all research databases were emailed (using the text in table 2 of the supplementary information) and asked to disclose the number of access requests and publications. Responses to this initial email were used to complete or update fields in the Access Database. Second emails were sent 5 weeks later to those who had not responded to the first email. A third and final email was sent a further 6 weeks later (11 weeks after the initial email) to those who had not responded to the first two emails. Emails were loaded into NVivo[8] and a content analysis conducted by two investigators who subsequently discussed and agreed on consensus categories.

Where conflicting information on a Research Database was noted from annual reports and subsequent email responses, the information from the email response was considered more up to date. The annual report template form was modified in 2011 adding a number of new

fields, although some researchers continued using the older version of the form after this date. Reports on the old form did not contain all the information required for this audit leading to missing categories for some research database records.

Results

453 Research Databases were initially identified, but then reduced to 354 eligible databases as illustrated in Figure 1. Welsh databases were excluded because only titles and reference numbers were included in HARP with no point of contact. Four duplicates were identified and excluded. 90 HARP entries were found to be renewals of previous applications, although these were difficult to initially identify as the titles and chief investigators were often not identical with the original studies. Many of these duplicates/renewals were only identified following email contact with researchers who complained they had received two emails for the same database. Once identified, all duplicate applications were combined, and renewal applications were combined with their parent (initial) application, but the start date of the initial application retained. The final list of 354 unique Research Databases had initial application dates ranging from May 2002 (when the first electronic records were compiled) to December 2017. The combination of data obtained from HARP and information obtained from annual and final reports was sufficient to fully populate the Access Database fields in 60 (17%) cases. Even following the three email contacts complete records were only obtained for 223 (63%) of the Research Databases. 44 (12%) invalid email addresses were identified following the first email to primary contacts, and when the second contact was subsequently used only 11 further responses were received. This left 33 (9%) databases that we were unable to contact. A few responses were received from individuals no longer involved with the Research Databases who provided updated contact details due to personnel changes.

Annual and Final Reports

The HRA stipulates that approved Research Databases submit annual reports for the duration that the database is collecting data and a final report if the database is closed[4]. Figure 2 shows a summary of annual reports that had been uploaded to HARP prior to contacting researchers by email. 54 (15%) Research Databases were less than a year old (meaning no annual report was yet due), and 108 (31%) had all or at least one annual report(s) on file. This left 192 (54%) of Research Databases with no annual report on file despite these being due (none filed 39% + received not filed 15%). 13 Research Databases dated prior to 2012 had no information on whether any reports had been received or requested. HARP did contain evidence (in the form of reminder letters held on file) that annual reports had on occasion been asked for, but such chasing emails/letters had not been sent or recorded in a systematic manner. Similarly there were 54 Research Databases where an annual report was noted as “Received Not Filed”. Here it seemed that although a letter was filed on HARP acknowledging receipt of an annual report, no electronic report

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

was present, although such reports may have been reviewed by the REC in hard copy but then not subsequently scanned and added to HARP. Of the 108 Research Databases with annual reports, only 32 (9% of all Research Databases in this study) were up to date with all reports.

Most Research Databases did not have completion dates and thus were open ended. Final reports were present for 16 (5%) Research Databases and, following email contact, a further 4 (1%) of Research Databases stated they had closed. It is impossible to determine how many of the 33 Research Databases without valid contact details were now closed and thus due a final report.

Amendments

Amendments are different from annual reports as they can be submitted at any time and normally notify changes of methodology or notification of significant event(s). 110 (31%) of Research Databases had at least one amendment recorded on HARP. Changes to database paperwork (such as version numbers, additional posters or advertising materials, changes of job title etc.) were the most common reason for an amendment with modifications of inclusion criteria, adding additional data linkages or including new participant groups, the next most common. Other less frequent amendments included changes in personnel, changes in process (different data capture methods or procedures), changes to location of the database and addition of new sites. No Research Databases reported any serious data breach.

Data Access Requests

The number of data access requests were known for 245 (69%) of the Research Databases. 123 of these (35% of total) reported no access requests, leaving all the access requests coming from only 122 (34% of total). Although the mean number of requests from these were 7.9, this was skewed by two outliers with 237 and 142 respectively. Of the 1948 total number of access requests, 1818 (93% of access requests) were granted. There were 52 requests noted as 'pending consideration' and two 'withdrawn'. As over 90% of access requests were granted overall, we considered the 'pending consideration' requests as granted, and the 'withdrawn' as not granted. Data summarising access requests and requests granted is presented in Figure 3.

Publications Resulting from Research Databases

The publication status was determined for 230 (65%) of the Research Databases. 'Publication' was defined to include presentations, conference abstracts and articles submitted for publication in professional journals. 88 (25%) reported no publication, with 142 (40%) declaring the 1868 publications. This gave a mean number of publications for all Research Databases with known publication status of 8.1, but this average is skewed by one major outlier with 315 publications, and a further two with over 80 publications. 31 (9% of

total) Research Databases had only one publication. A distribution of the number of publications coming from the research databases is shown in figure 4.

Age of Research Databases

Previous research looking at publication rates of projects reviewed by HRA RECs indicated that most projects take at least four years before a resulting publication in a peer reviewed journal is produced [7]. It might also be expected that the older a Research Database is the more likely it will be for other researchers to know about it and thus make a data access request. Here, the number of Research Databases approved per year is shown in figure 5, although it should be noted that some databases may have been in operation prior to the HRA application date. The mean age of all the Research Databases was 4.7 years, while the mean age of Research Databases with at least one publication was 5.8 years. Interestingly this compares to the mean age of a Research Database with at least one access request being 6.5 years. Figure 6 shows the total numbers of publications and access requests granted by the age of database.

Relationship between response to the Audit, data access, publication and age

MedCalc[9] was used to calculate odds ratios. There was a strong negative relationship between registration prior to 2012 and responding to the Audit (OR=0.52 p=0.005 CI 0.32-0.82). There was no significant relationship between age and publication status (OR 1.27, p = 0.27, CI 0.82-1.97). As previous evidence suggests publication becomes more likely after 4 years[7,10], we looked to see if a similar pattern emerged here by splitting the data into research databases younger and older than four years, but did not find any significant relationship (OR 1.28, p=0.26, CI 0.83-1.99). However, research databases with at least one data access request granted were significantly more likely to report at least one publication (OR 13.77, p<0.0001, CI 7.75-24.45). Out of the 354 Research Databases, 18 made some mention of patient or participant involvement (PPI) in their annual reports. This was strongly associated with having at least one publication or data request or both (OR 18.7, p<0.005, CI 2.46-142.12).

Observations from correspondence with Investigators

95 replies were received in response to our first email, 56 from the second, and a further 77 from the third. Responses often included comments explaining or further clarifying answers to the three questions we asked. A representative sample of responses are summarised in the supplementary information. Following coding, responses were grouped into two main categories: "Database access and sharing" and "Database management" as outlined in table 3 of the supplementary information.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Supérieur (ABES)

Discussion

The HRA Assessment Review Portal (HARP) is the authoritative database of all studies reviewed and given a favourable opinion by UK RECs. However, one important finding from this study was that the data contained in HARP was incomplete and in some cases inaccurate. The main reason for this was failure by researchers to send in required reports. There was also evidence that reports had been received in hard copy, perhaps viewed by the REC, but then not subsequently scanned and filed on HARP. While it must be acknowledged that the HRA can only populate HARP with the information it is given by researchers, this study seems to provide evidence to support the argument that more could be done by the HRA to ensure their records are complete and accurate. Information about data access requests received, granted, and publications relating to the database, could only be obtained for 60 (17%) of Research Databases based purely on the information in HARP, rising to 226 (64%) following email contact with the research teams. Concerningly we did not have valid contact details for 33 (almost 10%) of Research Databases, and although we gave up after three attempts, the HRA may need to follow these up further with the research sponsor. It was interesting, although perhaps to be expected, that the older databases were statistically less likely to reply to emails. Combining the 226 where we were able to obtain the necessary data items with the 33 that couldn't be contacted, we were still left with 95 (27%) databases where even following email contact not all the data we required was gathered.

As the concept of a Research Database includes storing and making data available for longer periods of time, it was not surprising that only a small number had provided final reports (indicating that the database was closed or closing). The email responses that we received indicated a number of reasons for closing databases including lack of funding, failure to gather the intended information, or changes in policies/legislation/clinical practice making the Research Database no longer relevant. However, none of these reasons would justify discarding data that could have potential future use, and although there are legal restrictions on the storage of identifiable patient information, there does seem to be an ethical argument for continued archiving of anonymised datasets either by sponsors or perhaps through other national or international arrangements. Further guidance from the HRA on what to do with "closed" Research Databases could be useful.

Despite Research Databases existing to store and share data, 67 (19% of total) reported that they had neither a publication nor allowed data access to other researchers. Of the rest 116 (33%) had granted access requests and (a mostly overlapping) 142 (40%) had produced publications (the discrepancy from 100% is due to having no information for 30%, and a smaller number with only partial information). The mean numbers of data access requests (7.9) and publications (8.1) per database (where these figures were known) could be viewed as indicating that the 30% or so of Research Databases that share data or publish are doing very well, however, these averages are distorted by a small number of very successful databases such as the *I-DSD (International Disorders of Sex Development) Research Database* with 237 granted access requests and 14 publications[11]. Similarly the *Searchable Online Database for MRC UK Brain Banks Network* reported 142 granted access requests and 315 publications. Another large research database, the *South London and*

Maudsley NHS Foundation Trust Biomedical Research Centre Case Register (SLaM BRC) [12] had 104 access requests granted, and although they named only a few publications, they did advise that an online search would undoubtedly find more. This suggests that for the larger databases the number of publications recorded here could be an underestimate. Interestingly, the features of these more successful Research Databases included long running support from large institutions and research councils, coupled with charity and institutional funding. They also seemed to show evidence of collaborative working with many contributing sites and participant involvement initiatives.

Calculating odds ratios did not demonstrate a link between age and data access or publications, but an increase in publications compared to access requests for databases aged between four and eight years (figure 6) support observations made from other types of studies [7,10,13] that it takes researchers about four or so years to obtain and analyse results, and then produce their first publication. However, there were fewer Research Database applications in the 2013 to 2015 period (figure 5), perhaps distorting our results. Odds ratios did, however, demonstrate a strong correlation between the granting of at least one access request and producing at least one publication (OR 13.77, $p<0.0001$, CI 7.75-24.45). Interestingly the average age of a database with one publication was 5.8 years, while the average for at least one granted access request was 6.5 years, indicating that many publications came from the database owners themselves. This may reflect the time taken to set up the database in the first place whereupon following the first publication other researchers become aware of the database and subsequently request access.

This fact that only 34% of research databases reported granting access requests, and 40% reported publications, is concerning ethically especially from the perspective of research participants who may have initially given consent for their data to be included in a research database with the belief that their data would be shared widely and thus be of use to multiple projects. Although the email responses from researchers did provide some valid reasons for not sharing data or publishing papers (for instance the research database being designed as part of a feasibility study, as prospective participant registry, or concerns around the possibility of re-identifying participants if the data was combined with other information held by third party researchers), more could be done to encourage researchers to at least acknowledge the database in their other work or publications[14], and thus remain accountable to the participants who contributed.

Analysis of the email responses also indicated a certain level of confusion over what constitutes a research database. In one case the researchers admitted that they had flagged their work as a research database in error, in another case an application was not renewed when the research team realised an ongoing favourable ethics opinion was not required for their specific type of study, and in other cases applications that had previously been flagged as another type of study were subsequently re-flagged as Research Databases or vice versa. One database reported they only chose to register as a Research Database to enable them to share information with a funder, and others admitted that they found it easier to apply as a Research Database rather than as a specific project so that they could share their data with collaborators and also use it for many different projects that they had not yet designed.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

Here the implication was that by calling their work a Research Database it would allow them more flexibility to use their own data.

Along with incorrect flagging, other reasons given for not sharing or publishing data included a lack of resources in terms of staffing or the funding required to promote the database as a resource. Here it was interesting to note that some of the Research Databases with the most access requests granted did charge to cover costs, and advertised these costs along with their access arrangements via their websites [11]. This seems like a good example of best practice that could be promoted more widely.

A number of studies justified the lack of access or publications by the amount of time required to gather enough data to make analysis worthwhile. Though this might be expected for databases within the first few years since application, some much older studies also used this excuse. This echoes evidence from elsewhere regarding no standard definition of what a reasonable time to prepare for data-sharing might be [15–17], although it may also be a consequence of some extremely long-running cohort studies being included in our sample.

One promising finding from this study was the high percentage of data access requests that were granted (93%). Here it was interesting to note that some databases reported screening requests or working with people wanting to make potential requests to ensure that the requests were suitable. Others reported lengthy application processes or publicising very specific approval criteria to try and reduce the number of rejected requests.

Study Limitations

The major limitation of this study was the incomplete records on the HARP database along with the absence of annual reports. Furthermore a pragmatic decision was made to limit the questions sent to researchers in our subsequent emails rather than send a more extensive survey or questionnaire. This resulted in often ambiguous replies from researchers making it difficult to complete all the fields in our Access database. An improvement to our methodology would therefore have involved sending a formal questionnaire or data entry form, perhaps similar to the templates produced by the HRA for final and annual reports. We also accepted a wide definition of the term “publication” to include peer reviewed publications, conference abstracts, posters and presentations. This was a potentially contentious decision as although peer reviewed research papers are the “gold standard” of scientific publication, there are a variety of other dissemination methods that are appropriate depending upon the situation[18]. For instance, the recent emphasis on “Patient and Public Involvement” (PPI) has tried to encourage researchers to produce bulletins and research summaries that are lay friendly and accessible[19]. While this should not be the only way research is disseminated, it is entirely valid for the purpose of maintaining accountability with research participants. It would perhaps therefore be a valuable future piece of work to determine what “appropriate” or “sufficient” publication/dissemination may look like for a Research Database. Interestingly the databases in our study that produced newsletters and bulletins as part of their PPI work were more likely to report publications and share their data with other researchers.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Acknowledgements

We thank Dr Janet Messer for support during this research and feedback on our initial report to the HRA.

Funding Statement

This work was funded by the Health Research Authority in the form of tuition fees and a small stipend for ST to complete a MRes under the supervision of SEK.

Competing Interests

SEK is chair of the Hampshire A HRA research ethics committee, the MOD research ethics committee, and a member of the HRA's Confidentiality Advisory Group (CAG). He is also an academic and the ethics advisor at the University of Portsmouth. ST is a lay member of the Hampshire B HRA research ethics committee. MB has no competing interests.

Author Contributions

SEK devised the project, arranged for funding and supervised ST. ST conducted the research and drafted the manuscript. MB supervised the qualitative elements of the research and coded the email responses in addition to ST. All authors contributed to the final manuscript.

Reporting Checklist

The SRQR (Standards for Reporting Qualitative Research) checklist was used to ensure accurate reporting.

Data Sharing

This study analysed data held in confidence by the UK Health Research Authority (HRA). Permission to access the original data can be requested by contacting the HRA.

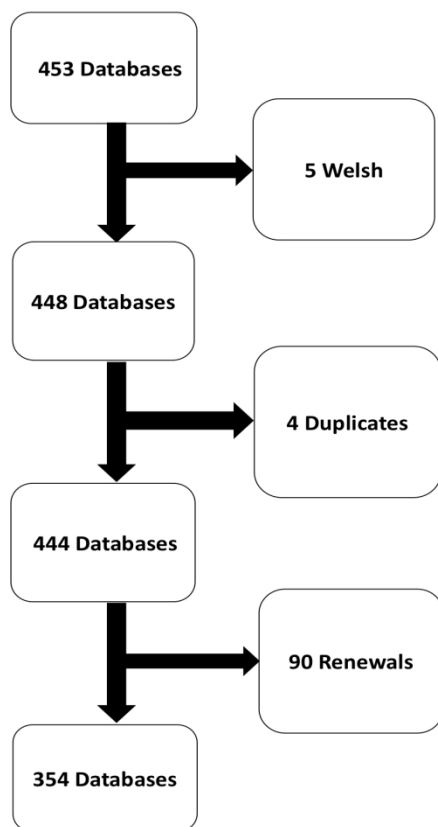
Patient and Public Involvement

As a government body the HRA involves patients and the public through their public involvement network. This research/audit was conducted as part of the HRA's transparency agenda to provide better information for patients and the public.

References

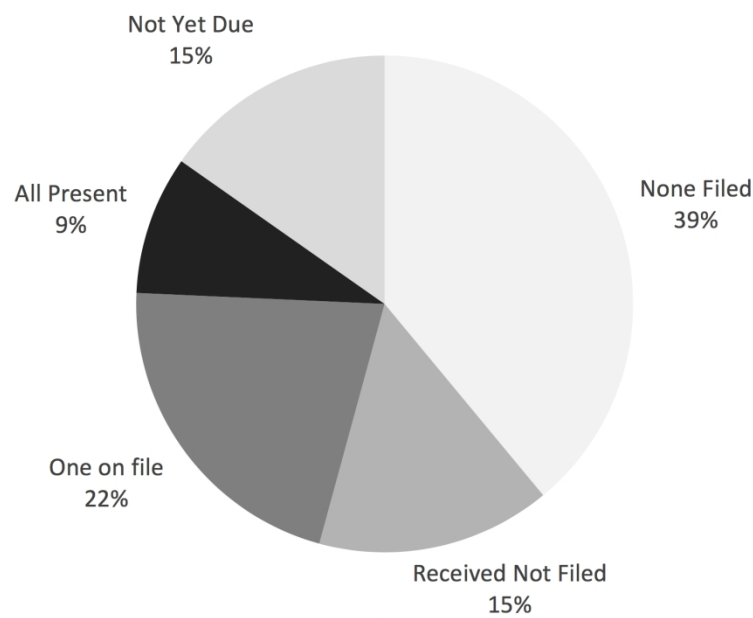
- 1 Kochhar S, Knoppers B, Gamble C, *et al*. Clinical trial data sharing: Here's the challenge. *BMJ Open* 2019;**9**:2–5. doi:10.1136/bmjopen-2019-032334
- 2 HRA. Research tissue banks and research databases. <https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/research-tissue-banks-and-research-databases/> (accessed 15 Apr2020).
- 3 HRA. What Approvals do I need? 2020. <https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/> (accessed 15 Apr2020).
- 4 HRA. Standard Operating Procedures. 2019.<https://www.hra.nhs.uk/about-us/committees-and-services/res-and-recs/research-ethics-committee-standard->

- operating-procedures/ (accessed 14 Apr2020).
- 5 AllTrials. What Does All Trials Registered and Reported Mean? AllTrials Website. 2020.<https://www.alltrials.net/find-out-more/all-trials/> (accessed 14 Apr2020).
 - 6 REWARD. REWARD. 2020.<http://rewardalliance.net/> (accessed 14 Apr2020).
 - 7 Begum R, Kolstoe S. Can UK NHS research ethics committees effectively monitor publication and outcome reporting bias? *BMC Med Ethics* 2015;**16**:51. doi:10.1186/s12910-015-0042-8
 - 8 QSR International Pty Ltd. NVivo qualitative data analysis software. 2018.
 - 9 Medcalc Software bvba. MedCalc. https://www.medcalc.org/calc/odds_ratio.php (accessed 14 Apr2020).
 - 10 Hopewell S, Clarke MJ, Stewart L, *et al*. Time to publication for results of clinical trials (Cochrane Review). *Cochrane Database Syst Rev* 2007;:MR000011. doi:10.1002/14651858.MR000011.pub2
 - 11 I-DSD. I-DSD Webpage. 2020.<https://home.i-dsd.org/> (accessed 14 Apr2020).
 - 12 Stewart R, Soremekun M, Perera G, *et al*. The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: Development and descriptive data. *BMC Psychiatry* 2009;**9**:1–12. doi:10.1186/1471-244X-9-51
 - 13 Suñé P, Suñé JM, Montoro JB. Positive Outcomes Influence the Rate and Time to Publication, but Not the Impact Factor of Publications of Clinical Trial Results. *PLoS One* 2013;**8**:1–8. doi:10.1371/journal.pone.0054583
 - 14 Kalager M, Adami H-O, Bretthauer M. Recognizing Data Generation. *N Engl J Med* 2016;**374**:1898–1898. doi:10.1056/NEJMc1603789
 - 15 Rani M, Buckley BS. Systematic archiving and access to health research data: rationale, current status and way forward. *Bull World Health Organ* 2012;**90**:932–9. doi:10.2471/BLT.12.105908
 - 16 Rathi V, Dzara K, Gross CP, *et al*. Sharing of clinical trial data among trialists: A cross sectional survey. *BMJ* 2012;**345**:1–10. doi:10.1136/bmj.e7570
 - 17 The International Consortium of Investigators for Fairness in Trial Data Sharing. Toward Fairness in Data Sharing. *N Engl J Med* 2016;**375**:405–7. doi:10.1056/NEJMp1605654
 - 18 McDonald S, Hopewell S, Clarke M, *et al*. Grey literature in meta-analyses of randomized trials of health care interventions (Review). *Cochrane Collab* 2008;:1–18. doi:10.1002/14651858.MR000010.pub3.www.cochranelibrary.com
 - 19 HRA. Informing Participants and Seeking Consent. <https://www.hra.nhs.uk/planning-and-improving-research/best-practice/informing-participants-and-seeking-consent/> (accessed 14 Apr2020).



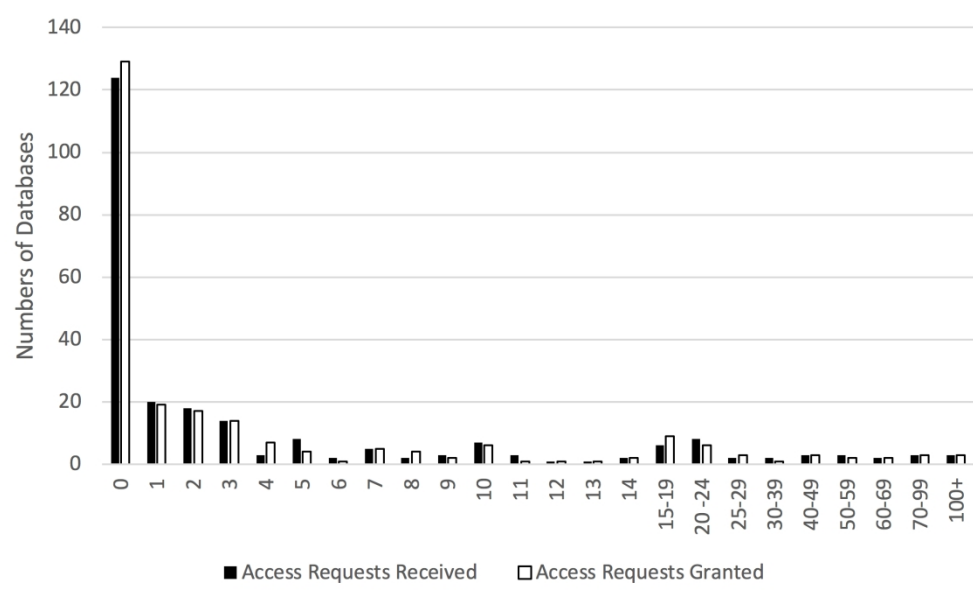
Inclusion flowchart

165x148mm (300 x 300 DPI)



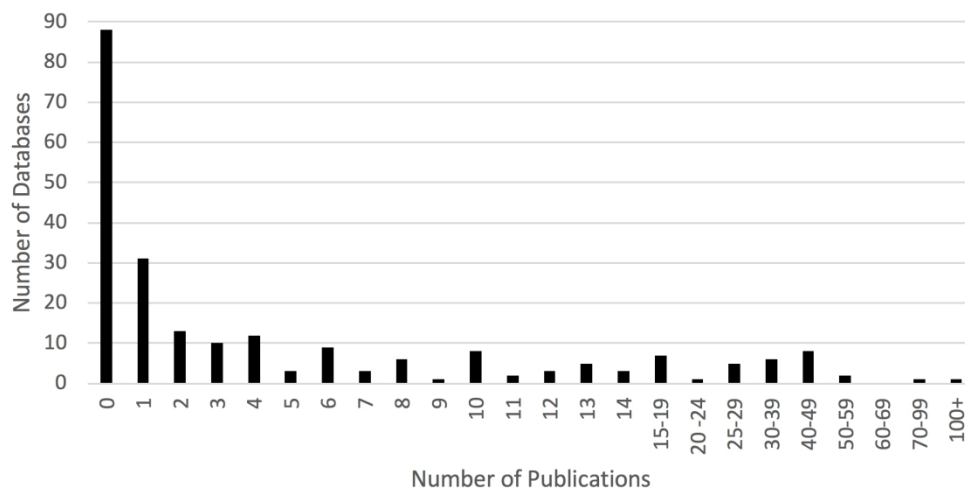
Annual reports contained on HARP

359x227mm (144 x 144 DPI)



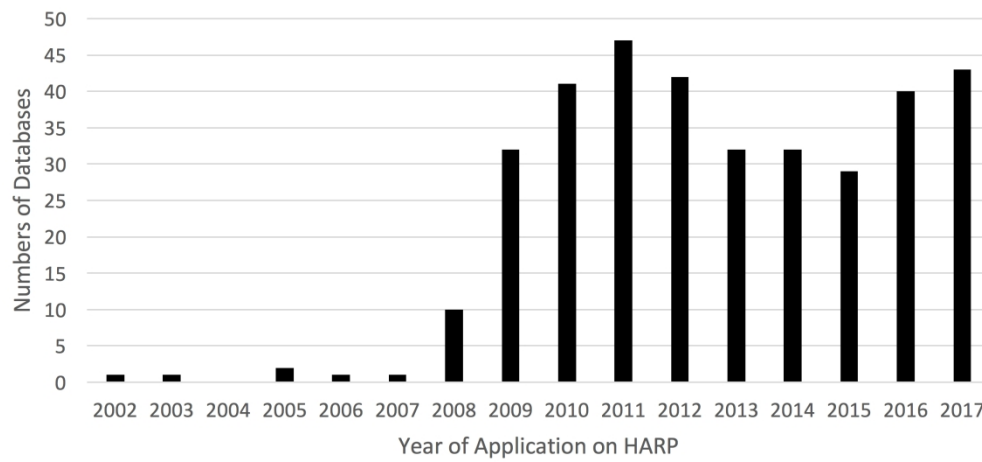
Data access requests received and granted grouped by frequency. NB, the “access requests granted” columns are sometimes higher than the “access requests received” columns because databases receiving multiple access requests did not always grant all of them.

390x232mm (144 x 144 DPI)



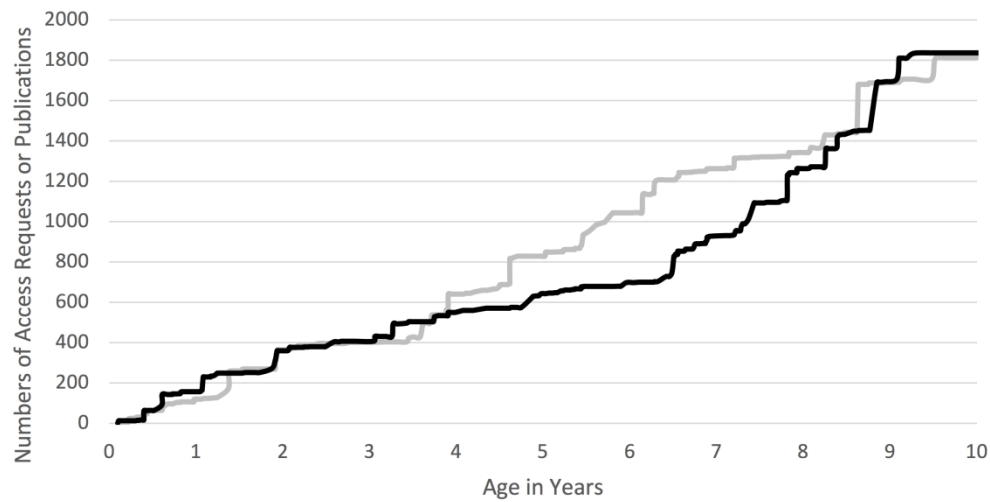
Distribution of publications coming from the research databases.

388x198mm (144 x 144 DPI)



Number and year of Research Database Applications.

470x224mm (144 x 144 DPI)



Cumulative numbers of publications and access requests granted for all databases by age of databases (e.g. at 5 years there are 641 access requests and 828 publications for all databases 5 years old and younger).

472x241mm (144 x 144 DPI)

Audit of data sharing and publications coming from research databases given a favourable opinion by UK research ethics committees

Samantha Trace, Mike Bracher and Simon Erik Kolstoe

Supplementary Information

1. Date Database Registered	9. Number of access requests granted
2. Database Title	10. Number of access requests received
3. Reviewing REC	11. Number of publications
4. Lead Applicant Name	12. Number of amendments
5. 1 st contact email address	13. Public Engagement Details
6. 2 nd contact email address	14. Response to email contact
7. Database Aims	15. Final report received (Y/N)
8. Annual Reports Received	

Table 1: Access Database fields

Dear <<Name of Lead Applicant>>
Re: <<Title of Research Database>>
The Health Research Authority is performing an audit on the usage of all Research Databases that have previously been granted a favourable ethics opinion.
We remind you that the continuing favourable ethics opinion (if relevant) is contingent upon regular submissions of information to the HRA through annual reports, until closure and submission of the final report.
As a point of contact for <<Title of Research Database>> please could you provide us with the following information about this database,

1. How many requests are received from researchers wishing to access the database per year?
2. How many requests for database access are granted per year?
3. Please list any publications that have resulted from research using the database?

We look forward to hearing from you at your earliest convenience

Table 2: Email sent out to contact email addresses

Enseignement Supérieur (ABES) .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Name	Description	References
<i>Database access and sharing</i>	<i>Responses relating to access and data sharing processes.</i>	38
Different understandings of what 'data sharing' means	Responses here provide evidence of a range of understandings of what 'data sharing' is, and how it operates in relation to the respondent's database.	5
Formal and informal access processes affect data on access	Responses indicating that formal and informal processes exist relating to access (typically informal approaches and formal applications). Many of these illustrate the lack of data on the former due to records only being available for the latter.	7
No external access	Responses here indicate that no external access has been granted to the database. These responses break down into two general types; those indicating explicitly that access was not facilitated; and those indicating only internal use of the database by the host organisation or research team.	26
Access Not Facilitated	Responses indicating explicitly that access was not facilitated to the database.	11
Internal Use Of Database Only	Responses indicating that the database was only used by the host organisation or research team.	15
<i>Database management</i>	<i>Responses relating to management of the database.</i>	10
Governance	Responses describing how the management and/or operation of the database has changed in response to changes in governance processes over time.	5
Information Technology (IT)	Responses describing how IT issues have affected operation and/or development of the database.	3
Workforce	Responses indicating workforce issues affecting the operation and/or viability of the database (e.g. availability of key staff).	2

No.	Topic	Item
Title and abstract		
S1	Title	Concise description of the nature and topic of the study Identifying the study as qualitative or indicating the approach (e.g., ethnography, grounded theory) or data collection methods (e.g., interview, focus group) is recommended
S2	Abstract	Summary of key elements of the study using the abstract format of the intended publication; typically includes background, purpose, methods, results, and conclusions
Introduction		
S3	Problem formulation	Description and significance of the problem/phenomenon studied; review of relevant theory and empirical work; problem statement
S4	Purpose or research question	Purpose of the study and specific objectives or questions
Methods		
S5	Qualitative approach and research paradigm	Qualitative approach (e.g., ethnography, grounded theory, case study, phenomenology, narrative research) and guiding theory if appropriate; identifying the research paradigm (e.g., postpositivist, constructivist/interpretivist) is also recommended; rationale ^b
S6	Researcher characteristics and reflexivity	Researchers’ characteristics that may influence the research, including personal attributes, qualifications/experience, relationship with participants, assumptions, and/or presuppositions; potential or actual interaction between researchers’ characteristics and the research questions, approach, methods, results, and/or transferability
S7	Context	Setting/site and salient contextual factors; rationale ^b
S8	Sampling strategy	How and why research participants, documents, or events were selected; criteria for deciding when no further sampling was necessary (e.g., sampling saturation); rationale ^b
S9	Ethical issues pertaining to human subjects	Documentation of approval by an appropriate ethics review board and participant consent, or explanation for lack thereof; other confidentiality and data security issues
S10	Data collection methods	Types of data collected; details of data collection procedures including (as appropriate) start and stop dates of data collection and analysis, iterative process, triangulation of sources/methods, and modification of procedures in response to evolving study findings; rationale ^b
S11	Data collection instruments and technologies	Description of instruments (e.g., interview guides, questionnaires) and devices (e.g., audio recorders) used for data collection; if/how the instrument(s) changed over the course of the study
S12	Units of study	Number and relevant characteristics of participants, documents, or events included in the study; level of participation (could be reported in results)
S13	Data processing	Methods for processing data prior to and during analysis, including transcription, data entry, data management and security, verification of data integrity, data coding, and anonymization/deidentification of excerpts
S14	Data analysis	Process by which inferences, themes, etc., were identified and developed, including the researchers involved in data analysis; usually references a specific paradigm or approach; rationale ^b
S15	Techniques to enhance trustworthiness	Techniques to enhance trustworthiness and credibility of data analysis (e.g., member checking, audit trail, triangulation); rationale ^b
Results/findings		
S16	Synthesis and interpretation	Main findings (e.g., interpretations, inferences, and themes); might include development of a theory or model, or integration with prior research or theory
S17	Links to empirical data	Evidence (e.g., quotes, field notes, text excerpts, photographs) to substantiate analytic findings
Discussion		
S18	Integration with prior work, implications, transferability, and contribution(s) to the field	Short summary of main findings; explanation of how findings and conclusions connect to, support, elaborate on, or challenge conclusions of earlier scholarship; discussion of scope of application/generalizability; identification of unique contribution(s) to scholarship in a discipline or field
S19	Limitations	Trustworthiness and limitations of findings
Other		
S20	Conflicts of interest	Potential sources of influence or perceived influence on study conduct and conclusions; how these were managed
S21	Funding	Sources of funding and other support; role of funders in data collection, interpretation, and reporting

^aThe authors created the SRQR by searching the literature to identify guidelines, reporting standards, and critical appraisal criteria for qualitative research; reviewing the reference lists of retrieved sources; and contacting experts to gain feedback. The SRQR aims to improve the transparency of all aspects of qualitative research by providing clear standards for reporting qualitative research.

^bThe rationale should briefly discuss the justification for choosing that theory, approach, method, or technique rather than other options available, the assumptions and limitations implicit in those choices, and how those choices influence study conclusions and transferability. As appropriate, the rationale for several items might be discussed together.

BMJ Open

Determining data sharing and number of publications coming from research databases that have been given a favourable opinion by UK research ethics committees

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-039756.R1
Article Type:	Original research
Date Submitted by the Author:	29-Jul-2020
Complete List of Authors:	Trace, Samantha; University of Portsmouth, School of Healthcare Professions Bracher, Mike; University of Southampton, Health Sciences Kolstoe, Simon; University of Portsmouth, School of Healthcare Professions
Primary Subject Heading:	Ethics
Secondary Subject Heading:	Research methods
Keywords:	QUALITATIVE RESEARCH, ETHICS (see Medical Ethics), AUDIT, Information management < BIOTECHNOLOGY & BIOINFORMATICS, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Supérieur (ABES).

Determining data sharing and number of publications coming from research databases that have been given a favourable opinion by UK research ethics committees

Samantha Trace¹, Mike Bracher² and Simon Kolstoe^{1,3}

¹ School of Health Care Professions, University of Portsmouth, PO1 2UP, UK

² School of Health Sciences, University of Southampton, SO17 1EN, UK

³ Correspondence to: Dr Simon Kolstoe; simon.kolstoe@port.ac.uk

Abstract

Objective To determine data sharing and number of publications coming from research databases that have been given a favourable opinion by UK NHS research ethics committees.

Design Cohort study

Inclusion Criteria & Setting All research databases listed on the UK Health Research Authority's Assessment Review Portal (HARP) that had received a favourable ethics opinion as of January 2018.

Main Outcome Measures Publications and data access requests either listed on HARP or notified through subsequent email correspondence.

Results Out of 354 eligible databases, 34% had granted access requests and 40% had produced at least one peer reviewed paper or conference abstract/talk. We could not establish contact with 9% of databases, and 19% reported no access requests or publications. Only 9% of databases were up to date with all annual reports. Email responses from database owners showed a range of attitudes towards data sharing.

Conclusion Less than half of Research Databases that have received a favourable opinion from NHS research ethics committees share their data and produce publications. There is also considerable variability in the operation of Research Databases and understanding of the purpose of Research Databases. This work was hampered by incomplete records due to researchers not submitting annual reports.

Key words

Research Database; Research Ethics; Data Sharing; Publication;

Word Count:

3,861

Article Summary

Strengths and limitations of this study

- By using the UK Health Research Authority’s “HARP” database we were able to identify all research databases using National Health Service (NHS) data that were registered as of January 2018
- We were able to identify both publications and access requests from the majority of databases
- We identified numerous incomplete records in the HARP database
- Research teams were not consistent in their definition of a research database, and it is likely that many relevant databases may not be registered with the HRA

Introduction

As data analysis processes continue to evolve, research databases represent increasingly important resources within healthcare research, yet there is evidence that they are currently under-utilised[1]. In the UK a research database is defined as:

“...a structured collection of individual-level personal information, which is stored for potential research purposes beyond the life of a specific research project with defined endpoints. Research purposes in this context refers to analysis of data to answer research questions in multiple projects.”[2]

The Health Research Authority (HRA) is the administrative body that convenes and organises research ethics committees (RECs) authorised to review studies involving human participants that take place within the National Health Service (NHS), as well as falling under certain legislation[3]. Although most of the HRA’s functions applies to research undertaken in England, its role coordinating policy, and managing the Integrated Research Application System (IRAS), gives it close links to the other devolved UK nations (Scotland, Wales and Northern Ireland) including access to records for audit and service improvement purposes. Through IRAS, the HRA flags research database applications, and provides a specific question set for researchers wishing to have their arrangements for collection, storage and use of data reviewed (including arrangements for release of non-identifiable data for analysis by external researchers). This requirement is outlined in the UK wide GAfREC (Governance Arrangements for Research Ethics Committees) policy whereby the:

“...collection of personal information from past or present users of health or social care services, or use of previously collected information from which individual users of these services could be identified, either directly from that information or from its combination with other information in, or likely to come into, the possession of someone to whom the information is made available”

always requires an ethics review, however the review of more generalised database projects by ethics committees:

... may have benefits by facilitating programmes of research using information on human subjects without a need for specific project-based applications. Applicants

1 *may seek generic ethical approval extending to specific projects undertaken using the*
2 *data, subject to conditions agreed with the REC.”[4]*
3
4

5 Consequently, in the UK, research databases differ from other types of research projects in
6 that they are normally intended to be used multiple times, over a longer period of time, and
7 perhaps by different research teams wishing to test a variety of hypotheses. When
8 reviewing research databases RECs therefore consider the access arrangements being made
9 available to researchers wishing to interrogate the database, including arrangements for
10 subsequent publication of research results. Indeed, this means there is an implicit
11 assumption that research databases will be used to generate many more publications than a
12 normal research project.
13
14
15

16 In order to test this assumption, and to benchmark UK performance with other national
17 studies[1], the HRA invited us to audit UK research database applications made through the
18 IRAS system. This request formed part of the wider “Transparency Agenda” being pursued
19 as a statutory duty by the HRA, but further encouraged by organisations such as the AllTrials
20 campaign[5] and the REWARD Alliance[6]. A previous audit by the HRA showed that only
21 one third of regular projects reviewed by RECs publish their results[7], raising a subsequent
22 concern that research database projects may also be underperforming in terms of
23 publishing outputs.
24
25
26

27 **Methods**

28
29 The initial inclusion criteria for this audit was projects flagged as Research Databases on the
30 HRA Assessment Review Portal (HARP) as of 1st January 2018. The number of eligible
31 databases were then reduced using the following criteria:
32
33

- 34 • Favourable Ethics opinion
- 35 • Not a duplicate record or renewal request
- 36 • Not a Welsh application
- 37
- 38
- 39

40 A Microsoft Access database was created with an entry for each research database and
41 information contained on HARP along with any uploaded annual or final reports were used
42 to populate the database fields listed in Table 1 of the supplementary information.
43 Following creation of the Access Database, primary contacts for all research databases were
44 emailed (using the text in table 2 of the supplementary information) and asked to disclose
45 the number of access requests and publications. Responses to this initial email were used to
46 complete or update fields in the Access Database. Second emails were sent 5 weeks later to
47 those who had not responded to the first email. A third and final email was sent a further 6
48 weeks later (11 weeks after the initial email) to those who had not responded to the first
49 two emails. Emails were loaded into NVivo[8] and a content analysis conducted by two
50 investigators who subsequently discussed and agreed on consensus categories.
51
52
53
54

55 Where conflicting information on a Research Database was noted from annual reports and
56 subsequent email responses, the information from the email response was considered more
57 up to date. The annual report template form was modified in 2011 adding a number of new
58 fields, although some researchers continued using the older version of the form after this
59
60

date. Reports on the old form did not contain all the information required for this audit leading to missing categories for some research database records.

Results

453 Research Databases were initially identified, but then reduced to 354 eligible databases after excluding four duplicates, Welsh databases (because only titles and reference numbers were included in HARP with no point of contact), and 90 HARP entries that were renewals of previous applications. These latter entries were difficult to initially identify as the titles and chief investigators were often not identical with the original studies. Indeed many of these duplicates/renewals were only identified following email contact with researchers who complained they had received two emails for the same database. Once identified, all duplicate applications were combined, and renewal applications were combined with their parent (initial) application, but the start date of the initial application retained. The final list of 354 unique Research Databases had initial application dates ranging from May 2002 (when the first electronic records were compiled) to December 2017. The combination of data obtained from HARP and information obtained from annual and final reports was sufficient to fully populate the Access Database fields in 60 (17%) cases. Even following the three email contacts complete records were only obtained for 223 (63%) of the Research Databases. 44 (12%) invalid email addresses were identified following the first email to primary contacts, and when the second contact was subsequently used only 11 further responses were received. This left 33 (9%) databases that we were unable to contact. A few responses were received from individuals no longer involved with the Research Databases who provided updated contact details due to personnel changes.

Annual and Final Reports

The HRA stipulates that approved Research Databases submit annual reports for the duration that the database is collecting data and a final report if the database is closed[4]. Figure 1 shows a summary of annual reports that had been uploaded to HARP prior to contacting researchers by email. 54 (15%) Research Databases were less than a year old (meaning no annual report was yet due), and 108 (31%) had all or at least one annual report(s) on file. This left 192 (54%) of Research Databases with no annual report on file despite these being due (none filed 39% + received not filed 15%). 13 Research Databases dated prior to 2012 had no information on whether any reports had been received or requested. HARP did contain evidence (in the form of reminder letters held on file) that annual reports had on occasion been asked for, but such chasing emails/letters had not been sent or recorded in a systematic manner. Similarly there were 54 Research Databases where an annual report was noted as “Received Not Filed”. Here it seemed that although a letter was filed on HARP acknowledging receipt of an annual report, no electronic report was present, although such reports may have been reviewed by the REC in hard copy but then not subsequently scanned and added to HARP. Of the 108 Research Databases with

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

annual reports, only 32 (9% of all Research Databases in this study) were up to date with all reports.

Most Research Databases did not have completion dates and thus were open ended. Final reports were present for 16 (5%) Research Databases and, following email contact, a further 4 (1%) of Research Databases stated they had closed. It is impossible to determine how many of the 33 Research Databases without valid contact details were now closed and thus due a final report.

Amendments

Amendments are different from annual reports as they can be submitted at any time and normally notify changes of methodology or notification of significant event(s). 110 (31%) of Research Databases had at least one amendment recorded on HARP. Changes to database paperwork (such as version numbers, additional posters or advertising materials, changes of job title etc.) were the most common reason for an amendment with modifications of inclusion criteria, adding additional data linkages or including new participant groups, the next most common. Other less frequent amendments included changes in personnel, changes in process (different data capture methods or procedures), changes to location of the database and addition of new sites. No Research Databases reported any serious data breach.

Data Access Requests

The number of data access requests were known for 245 (69%) of the Research Databases. 123 of these (35% of total) reported no access requests, leaving all the access requests coming from only 122 (34% of total). Although the mean number of requests from these were 7.9, this was skewed by two outliers with 237 and 142 respectively. Of the 1948 total number of access requests, 1818 (93% of access requests) were granted. There were 52 requests noted as 'pending consideration' and two 'withdrawn'. As over 90% of access requests were granted overall, we considered the 'pending consideration' requests as granted, and the 'withdrawn' as not granted. Data summarising access requests and requests granted is presented in Figure 2.

Publications Resulting from Research Databases

The publication status was determined for 230 (65%) of the Research Databases. 'Publication' was defined to include presentations, conference abstracts and articles submitted for publication in professional journals. 88 (25%) reported no publication, with 142 (40%) declaring the 1868 publications. This gave a mean number of publications for all Research Databases with known publication status of 8.1, but this average is skewed by one major outlier with 315 publications, and a further two with over 80 publications. 31 (9% of total) Research Databases had only one publication. A distribution of the number of publications coming from the research databases is shown in figure 3.

Age of Research Databases

Previous research looking at publication rates of projects reviewed by HRA RECs indicated that most projects take at least four years before a resulting publication in a peer reviewed journal is produced [7]. It might also be expected that the older a Research Database is the more likely it will be for other researchers to know about it and thus make a data access request. Here, the number of Research Databases approved per year is shown in figure 4, although it should be noted that some databases may have been in operation prior to the HRA application date. The mean age of all the Research Databases was 4.7 years, while the mean age of Research Databases with at least one publication was 5.8 years. Interestingly this compares to the mean age of a Research Database with at least one access request being 6.5 years. Figure 5 shows the total numbers of publications and access requests granted by the age of database.

Relationship between response to the Audit, data access, publication and age

MedCalc[9] was used to calculate odds ratios. There was a strong negative relationship between registration prior to 2012 and responding to the Audit (OR=0.52 p=0.005 CI 0.32-0.82). There was no significant relationship between age and publication status (OR 1.27, p = 0.27, CI 0.82-1.97). As previous evidence suggests publication becomes more likely after 4 years[7,10], we looked to see if a similar pattern emerged here by splitting the data into research databases younger and older than four years, but did not find any significant relationship (OR 1.28, p=0.26, CI 0.83-1.99). However, research databases with at least one data access request granted were significantly more likely to report at least one publication (OR 13.77, p<0.0001, CI 7.75-24.45). Out of the 354 Research Databases, 18 made some mention of patient or participant involvement (PPI) in their annual reports. This was strongly associated with having at least one publication or data request or both (OR 18.7, p<0.005, CI 2.46-142.12).

Observations from correspondence with Investigators

95 replies were received in response to our first email, 56 from the second, and a further 77 from the third. Responses often included comments explaining or further clarifying answers to the three questions we asked. A representative sample of responses are summarised in the supplementary information. Following coding, responses were grouped into two main categories: "Database access and sharing" and "Database management" as outlined in table 3 of the supplementary information.

Discussion

The HRA Assessment Review Portal (HARP) is the authoritative database of all studies reviewed and given a favourable opinion by UK RECs. However, one important finding from this study was that the data contained in HARP was incomplete and in some cases inaccurate. The main reason for this was failure by researchers to send in required reports.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

There was also evidence that reports had been received in hard copy, perhaps viewed by the REC, but then not subsequently scanned and filed on HARP. While it must be acknowledged that the HRA can only populate HARP with the information it is given by researchers, this study seems to provide evidence to support the argument that more could be done by the HRA to ensure their records are complete and accurate. Information about data access requests received, granted, and publications relating to the database, could only be obtained for 60 (17%) of Research Databases based purely on the information in HARP, rising to 226 (64%) following email contact with the research teams. Concerningly we did not have valid contact details for 33 (almost 10%) of Research Databases, and although we gave up after three attempts, the HRA may need to follow these up further with the research sponsor. It was interesting, although perhaps to be expected, that the older databases were statistically less likely to reply to emails. Combining the 226 where we were able to obtain the necessary data items with the 33 that couldn't be contacted, we were still left with 95 (27%) databases where even following email contact not all the data we required was gathered.

As the concept of a Research Database includes storing and making data available for longer periods of time, it was not surprising that only a small number had provided final reports (indicating that the database was closed or closing). The email responses that we received indicated a number of reasons for closing databases including lack of funding, failure to gather the intended information, or changes in policies/legislation/clinical practice making the Research Database no longer relevant. However, while there are legal restrictions on the storage of identifiable patient information (through legislation such as the European General Data Protection Regulation), concerns regarding reproducibility and the importance of "Open data"[11] mean that archiving of anonymised datasets either by sponsors or perhaps through other national or international arrangements is increasingly becoming expected. Further guidance from the HRA on what to do with "closed" Research Databases could be useful.

Despite Research Databases existing to store and share data, 67 (19% of total) reported that they had neither a publication nor allowed data access to other researchers. Of the rest 116 (33%) had granted access requests and (a mostly overlapping) 142 (40%) had produced publications (the discrepancy from 100% is due to having no information for 30%, and a smaller number with only partial information). The mean numbers of data access requests (7.9) and publications (8.1) per database (where these figures were known) could be viewed as indicating that the 30% or so of Research Databases that share data or publish are doing very well, however, these averages are distorted by a small number of very successful databases such as the *I-DSD (International Disorders of Sex Development) Research Database* with 237 granted access requests and 14 publications[12]. Similarly the *Searchable Online Database for MRC UK Brain Banks Network* reported 142 granted access requests and 315 publications. Another large research database, the *South London and Maudsley NHS Foundation Trust Biomedical Research Centre Case Register (SLaM BRC)* [13] had 104 access requests granted, and although they named only a few publications, they did advise that an online search would undoubtedly find more. This suggests that for the larger databases the number of publications recorded here could be an underestimate. Interestingly, the features of these more successful Research Databases included long

running support from large institutions and research councils, coupled with charity and institutional funding. They also seemed to show evidence of collaborative working with many contributing sites and participant involvement initiatives.

Calculating odds ratios did not demonstrate a link between age and data access or publications, but an increase in publications compared to access requests for databases aged between four and eight years (figure 5) support observations made from other types of studies [7,10,14] that it takes researchers about four or so years to obtain and analyse results, and then produce their first publication. However, there were fewer Research Database applications in the 2013 to 2015 period (figure 4), perhaps distorting our results. Odds ratios did, however, demonstrate a strong correlation between the granting of at least one access request and producing at least one publication (OR 13.77, $p<0.0001$, CI 7.75-24.45). Interestingly the average age of a database with one publication was 5.8 years, while the average for at least one granted access request was 6.5 years, indicating that many publications came from the database owners themselves. This may reflect the time taken to set up the database in the first place whereupon following the first publication other researchers become aware of the database and subsequently request access.

This fact that only 34% of research databases reported granting access requests, and 40% reported publications, is concerning ethically especially from the perspective of research participants who may have initially given consent for their data to be included in a research database with the belief that their data would be shared widely and thus be of use to multiple projects. Although the email responses from researchers did provide some valid reasons for not sharing data or publishing papers (for instance the research database being designed as part of a feasibility study, as prospective participant registry, or concerns around the possibility of re-identifying participants if the data was combined with other information held by third party researchers), more could be done to encourage researchers to at least acknowledge the database in their other work or publications[15], and thus remain accountable to the participants who contributed.

Analysis of the email responses also indicated a certain level of confusion over what constitutes a research database. In one case the researchers admitted that they had flagged their work as a research database in error, in another case an application was not renewed when the research team realised an ongoing favourable ethics opinion was not required for their specific type of study, and in other cases applications that had previously been flagged as another type of study were subsequently re-flagged as Research Databases or vice versa. One database reported they only chose to register as a Research Database to enable them to share information with a funder, and others admitted that they found it easier to apply as a Research Database rather than as a specific project so that they could share their data with collaborators and also use it for many different projects that they had not yet designed. Here the implication was that by calling their work a Research Database it would allow them more flexibility to use their own data.

Along with incorrect flagging, other reasons given for not sharing or publishing data included a lack of resources in terms of staffing or the funding required to promote the database as a resource. Here it was interesting to note that some of the Research

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Supérieur (ABES)

Databases with the most access requests granted did charge to cover costs, and advertised these costs along with their access arrangements via their websites [12].

A number of studies justified the lack of access or publications by the amount of time required to gather enough data to make analysis worthwhile. Though this might be expected for databases within the first few years since application, some much older studies also used this excuse. This echoes evidence from elsewhere regarding no standard definition of what a reasonable time to prepare for data-sharing might be [16–18], although it may also be a consequence of some extremely long-running cohort studies being included in our sample.

One promising finding from this study was the high percentage of data access requests that were granted (93%). Here it was interesting to note that some databases reported screening requests or working with people wanting to make potential requests to ensure that the requests were suitable. Others reported lengthy application processes or publicising very specific approval criteria to try and reduce the number of rejected requests.

Study Limitations

The major limitation of this study was the incomplete records on the HARP database along with the absence of annual reports. Furthermore a pragmatic decision was made to limit the questions sent to researchers in our subsequent emails rather than send a more extensive survey or questionnaire. This resulted in often ambiguous replies from researchers making it difficult to complete all the fields in our Access database. An improvement to our methodology would therefore have involved sending a formal questionnaire or data entry form, perhaps similar to the templates produced by the HRA for final and annual reports.

We also only looked at studies that had been prospectively labelled as databases. It would be interesting to determine how many other types of studies subsequently decided to establish databases as part of their open access/data sharing arrangements. This would not be a trivial task as it would involve writing to all chief investigators registered on HARP, but would potentially identify further relevant databases.

We also accepted a wide definition of the term “publication” to include peer reviewed publications, conference abstracts, posters and presentations. This was a potentially contentious decision as although peer reviewed research papers are the “gold standard” of scientific publication, there are a variety of other dissemination methods that are appropriate depending upon the situation[19]. For instance, the recent emphasis on “Patient and Public Involvement” (PPI) has tried to encourage researchers to produce bulletins and research summaries that are lay friendly and accessible[20]. While this should not be the only way research is disseminated, it is entirely valid for the purpose of maintaining accountability with research participants. It would perhaps therefore be a valuable future piece of work to determine what “appropriate” or “sufficient” publication/dissemination may look like for a Research Database. Interestingly the databases in our study that produced newsletters and bulletins as part of their PPI work were more likely to report publications and share their data with other researchers.

Acknowledgements

We thank Dr Janet Messer for support during this research and feedback on our initial report to the HRA.

Funding Statement

This work was funded by the Health Research Authority (#15182) in the form of tuition fees and a small stipend for ST to complete a MRes under the supervision of SEK.

Competing Interests

SEK is chair of the Hampshire A HRA research ethics committee, the MOD research ethics committee, and a member of the HRA’s Confidentiality Advisory Group (CAG). He is also an academic and the ethics advisor at the University of Portsmouth. ST is a lay member of the Hampshire B HRA research ethics committee. MB has no competing interests.

Author Contributions

SEK devised the project, arranged for funding and supervised ST. ST conducted the research and drafted the manuscript. MB supervised the qualitative elements of the research and coded the email responses in addition to ST. All authors contributed to the final manuscript.

Reporting Checklist

The SRQR (Standards for Reporting Qualitative Research) checklist was used to ensure accurate reporting.

Data Sharing

This study analysed data held in confidence by the UK Health Research Authority (HRA). Permission to access the original data can be requested by contacting the authors or the HRA.

Patient and Public Involvement

As a government body the HRA involves patients and the public through their public involvement network. This research/audit was conducted as part of the HRA’s transparency agenda to provide better information for patients and the public.

Approvals

As an audit of existing records (secondary data) held by a regulator this work did not require an ethics review. Both SEK and ST have regular access to HARP records as part of their roles on HRA research ethics committees.

References

- 1 Kochhar S, Knoppers B, Gamble C, *et al.* Clinical trial data sharing: Here’s the challenge. *BMJ Open* 2019;**9**:2–5. doi:10.1136/bmjopen-2019-032334
- 2 Health Research Authority. Research tissue banks and research databases - Health Research Authority. <https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/research-tissue-banks-and-research->

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

- databases/ (accessed 14 May 2019).
- 3 Health Research Authority. What Approvals do I need? 2018.<https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/hra-approval/> (accessed 14 May 2019).
- 4 Health Research Authority. Standard Operating Procedures. 2017.<http://www.hra.nhs.uk/documents/2017/01/standard-operating-procedures-version-7-2.pdf> (accessed 14 May 2019).
- 5 AllTrials. All Trials Declaration. 2013.<http://www.alltrials.net/wp-content/uploads/2013/09/What-does-all-trials-registered-and-reported-mean.pdf> (accessed 14 May 2019).
- 6 REWARD Alliance. 2018.<http://rewardalliance.net/> (accessed 14 May 2019).
- 7 Begum R, Kolstoe S. Can UK NHS research ethics committees effectively monitor publication and outcome reporting bias? *Ethics in Biomedical Research. BMC Med Ethics* 2015;**16**:51. doi:10.1186/s12910-015-0042-8
- 8 QSR International Pty Ltd. NVivo qualitative data analysis software. 2018.<https://www.qsrinternational.com/nvivo/home>
- 9 Medcalc.org. MedCalc Software. https://www.medcalc.org/calc/odds_ratio.php (accessed 14 May 2019).
- 10 Hopewell S, Clarke MJ, Stewart L, *et al*. Time to publication for results of clinical trials. *Cochrane Database Syst. Rev.* 2007;:MR000011. doi:10.1002/14651858.MR000011.pub2
- 11 Munafò MR, Nosek BA, Bishop DVM, *et al*. A manifesto for reproducible science. *Nat Hum Behav* 2017;**1**:0021. doi:10.1038/s41562-016-0021
- 12 I-DSD. I-DSD Registry. 2018.<https://www.gla.ac.uk/schools/medicine/research/childhealth/researchinterests/i-dsdproject/i-dsdmembership/> (accessed 14 May 2019).
- 13 Stewart R, Soremekun M, Perera G, *et al*. The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: Development and descriptive data. *BMC Psychiatry* 2009;**9**:1–12. doi:10.1186/1471-244X-9-51
- 14 Suñé P, Suñé JM, Montoro JB. Positive Outcomes Influence the Rate and Time to Publication, but Not the Impact Factor of Publications of Clinical Trial Results. *PLoS One* 2013;**8**:1–8. doi:10.1371/journal.pone.0054583
- 15 Kalager M, Adami H-O, Bretthauer M. Recognizing Data Generation. *N Engl J Med* 2016;**374**:1898–1898. doi:10.1056/nejmc1603789
- 16 Rani M, Buckley BS. Systematic archiving and access to health research data:

rationale, current status and way forward. *Bull World Health Organ* 2016;**90**:932–9. doi:10.2471/blt.12.105908

17 Rath V, Dzara K, Gross CP, *et al.* Sharing of clinical trial data among trialists: A cross sectional survey. *BMJ* 2012;**345**:1–10. doi:10.1136/bmj.e7570

18 The International Consortium of Investigators for Fairness in Trial Data Sharing. Toward Fairness in Data Sharing. *N Engl J Med* 2016;**375**:405–7. doi:10.1056/nejmp1605654

19 Hopewell S, Mcdonald S, Clarke M, *et al.* Grey literature in meta-analyses of randomized trials of health care interventions (Review). *Library (Lond)* 2009;:1–18. doi:10.1002/14651858.MR000010.pub3.www.cochranelibrary.com

20 Health Research Authority. Consent and Participant Information Sheet. <http://www.hra-decisiontools.org.uk/consent/> (accessed 14 May 2019).

Figure Captions:

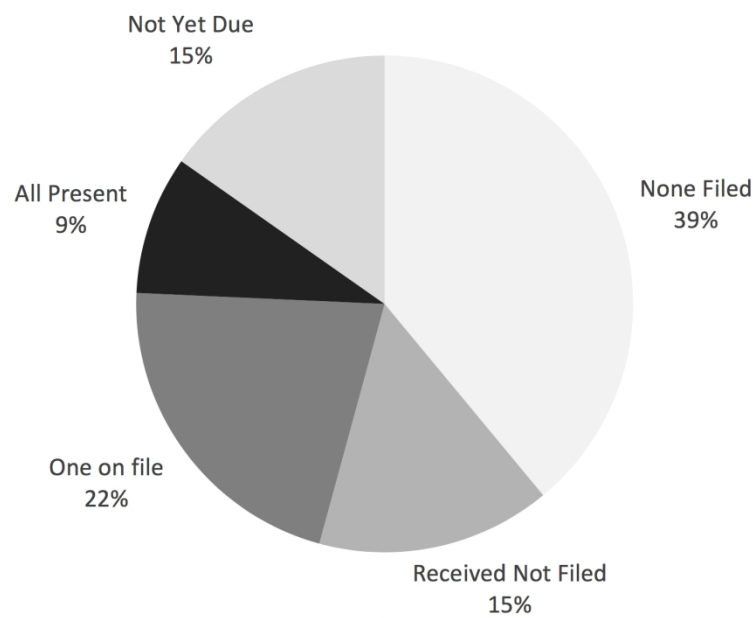
Figure 1: Annual reports contained on HARP

Figure 2: Data access requests received and granted grouped by frequency. NB, the “access requests granted” columns are sometimes higher than the “access requests received” columns because databases receiving multiple access requests did not always grant all of them.

Figure 3: Distribution of publications coming from the research databases.

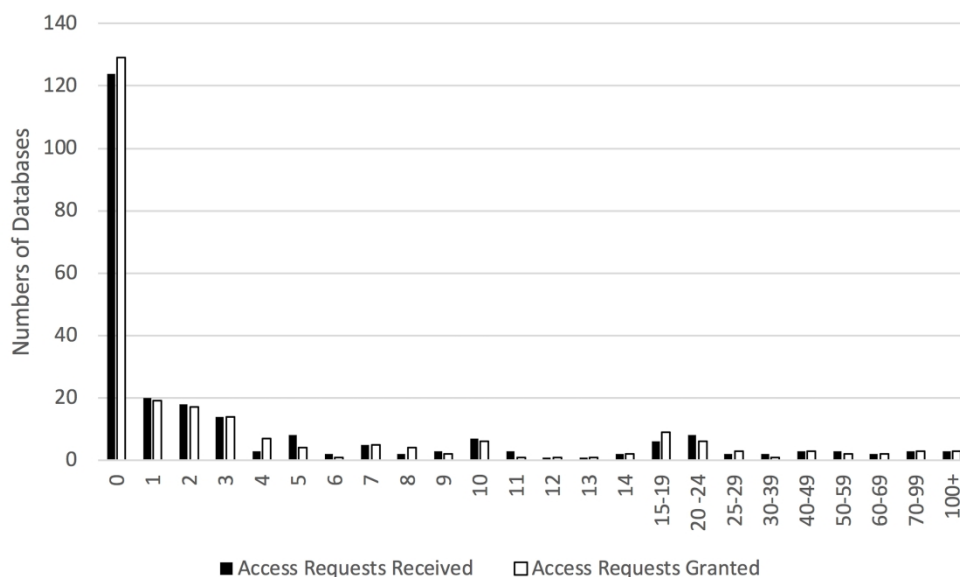
Figure 4: Number and year of Research Database Applications.

Figure 5: Cumulative numbers of publications (grey line) and access requests (black line) granted for all databases by age of databases (e.g. at 5 years there are 641 access requests and 828 publications for all databases 5 years old and younger).



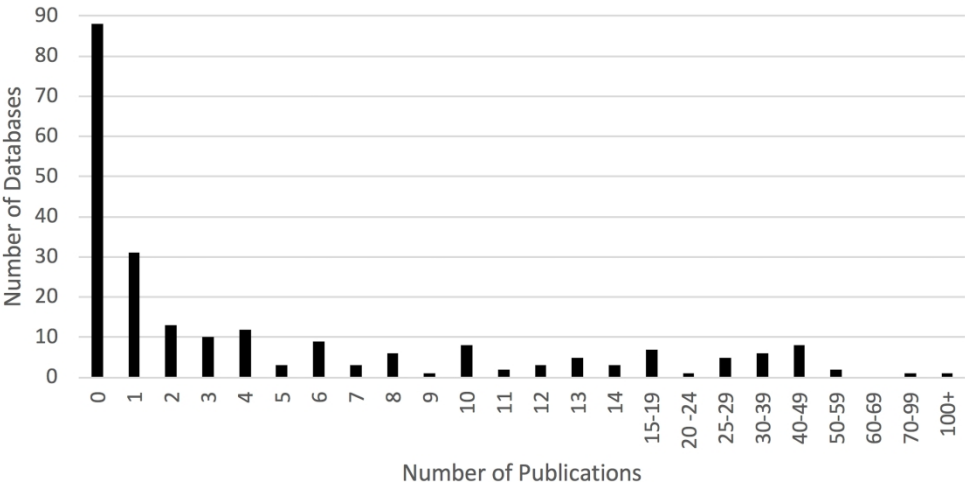
Annual reports contained on HARP

172x109mm (300 x 300 DPI)



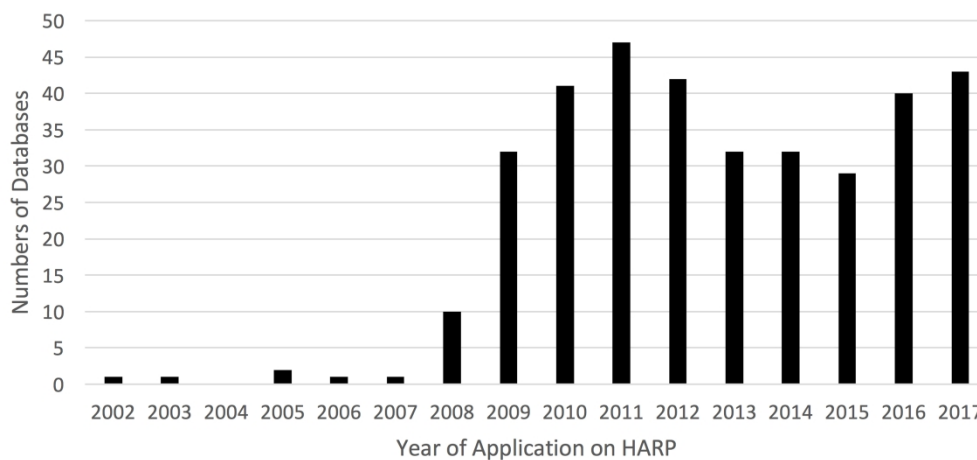
Data access requests received and granted grouped by frequency. NB, the “access requests granted” columns are sometimes higher than the “access requests received” columns because databases receiving multiple access requests did not always grant all of them.

390x232mm (300 x 300 DPI)



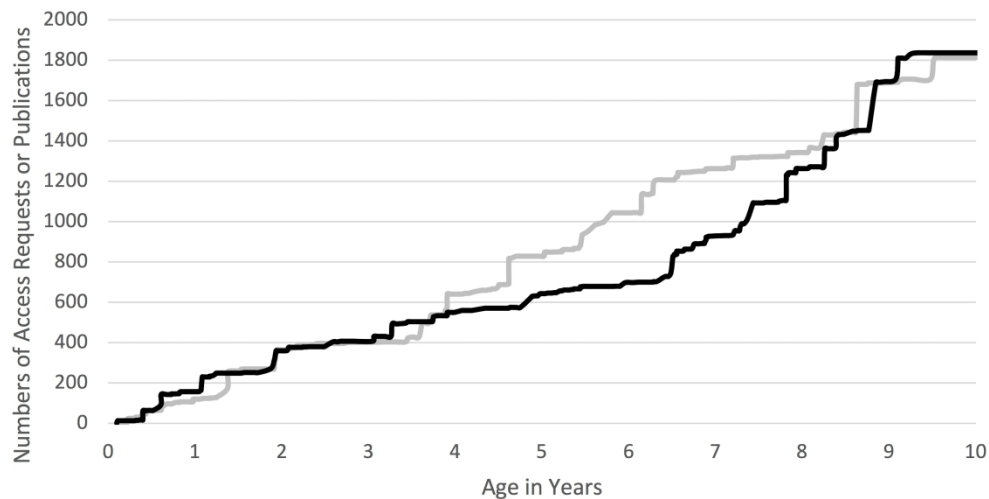
Distribution of publications coming from the research databases.

388x198mm (300 x 300 DPI)



Number and year of Research Database Applications.

471x224mm (300 x 300 DPI)



Cumulative numbers of publications (grey line) and access requests (black line) granted for all databases by age of databases (e.g. at 5 years there are 641 access requests and 828 publications for all databases 5 years old and younger).

472x241mm (300 x 300 DPI)

Audit of data sharing and publications coming from research databases given a favourable opinion by UK research ethics committees

Samantha Trace, Mike Bracher and Simon Erik Kolstoe

Supplementary Information

1. Date Database Registered	9. Number of access requests granted
2. Database Title	10. Number of access requests received
3. Reviewing REC	11. Number of publications
4. Lead Applicant Name	12. Number of amendments
5. 1 st contact email address	13. Public Engagement Details
6. 2 nd contact email address	14. Response to email contact
7. Database Aims	15. Final report received (Y/N)
8. Annual Reports Received	

Table 1: Access Database fields

Dear <<Name of Lead Applicant>>
Re: <<Title of Research Database>>
The Health Research Authority is performing an audit on the usage of all Research Databases that have previously been granted a favourable ethics opinion.
We remind you that the continuing favourable ethics opinion (if relevant) is contingent upon regular submissions of information to the HRA through annual reports, until closure and submission of the final report.
As a point of contact for <<Title of Research Database>> please could you provide us with the following information about this database,

1. How many requests are received from researchers wishing to access the database per year?
2. How many requests for database access are granted per year?
3. Please list any publications that have resulted from research using the database?

We look forward to hearing from you at your earliest convenience

Table 2: Email sent out to contact email addresses

Name	Description	References
<i>Database access and sharing</i>	<i>Responses relating to access and data sharing processes.</i>	<i>38</i>
Different understandings of what 'data sharing' means	Responses here provide evidence of a range of understandings of what 'data sharing' is, and how it operates in relation to the respondent's database.	5
Formal and informal access processes affect data on access	Responses indicating that formal and informal processes exist relating to access (typically informal approaches and formal applications). Many of these illustrate the lack of data on the former due to records only being available for the latter.	7
No external access	Responses here indicate that no external access has been granted to the database. These responses break down into two general types; those indicating explicitly that access was not facilitated; and those indicating only internal use of the database by the host organisation or research team.	26
Access Not Facilitated	Responses indicating explicitly that access was not facilitated to the database.	11
Internal Use Of Database Only	Responses indicating that the database was only used by the host organisation or research team.	15
<i>Database management</i>	<i>Responses relating to management of the database.</i>	<i>10</i>
Governance	Responses describing how the management and/or operation of the database has changed in response to changes in governance processes over time.	5
Information Technology (IT)	Responses describing how IT issues have affected operation and/or development of the database.	3
Workforce	Responses indicating workforce issues affecting the operation and/or viability of the database (e.g. availability of key staff).	2

Table 3: Coding of emails received