

# BMJ Open Data sharing: experience of accessing individual patient data from completed randomised controlled trials in vascular and cognitive medicine

Polly Scutt,<sup>1,2</sup> Lisa J Woodhouse ,<sup>1</sup> Alan A Montgomery,<sup>3</sup> Philip M Bath<sup>1</sup>

**To cite:** Scutt P, Woodhouse LJ, Montgomery AA, *et al*. Data sharing: experience of accessing individual patient data from completed randomised controlled trials in vascular and cognitive medicine. *BMJ Open* 2020;**10**:e038765. doi:10.1136/bmjopen-2020-038765

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-038765>).

PS and LJW are joint first authors.

Received 24 March 2020  
Revised 10 June 2020  
Accepted 01 July 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Stroke, Division of Clinical Neuroscience, University of Nottingham, Nottingham, UK  
<sup>2</sup>Hearing Sciences, NIHR Nottingham Biomedical Research Centre, Nottingham, UK

<sup>3</sup>Nottingham Clinical Trials Unit, University of Nottingham, Nottingham, UK

## Correspondence to

Philip M Bath;  
[philip.bath@nottingham.ac.uk](mailto:philip.bath@nottingham.ac.uk)

## ABSTRACT

**Objectives** Meta-analysis based on individual patient data (IPD) from randomised trials is superior to using published summary data since it facilitates subgroup and multiple variable analyses. Guidelines and funders expect that researchers share IPD for bona fide analyses, but in practice, this is done variably. Here, we report the experience of obtaining IPD for two collaborative analysis studies.

**Setting** Two linked studies required IPD from published randomised trials. The leading researchers for eligible trials were approached and asked to share IPD including trial characteristics, patient demographics, baseline clinical data and outcome measures.

**Participants** Participants in eligible randomised controlled trials included patients with or at risk of cognitive decline/vascular events.

**Primary and secondary outcome measures** Numbers (%) of trials where the leading researcher responded favourably/negatively or did not respond. If negative, reasons behind the response were collected. If positive, methods used to share IPD were recorded.

**Results** Across the two studies, 391 completed trials were identified. Email addresses for researchers were found for 313 (80%) of the trials. One hundred and forty-eight (47%) researchers did not respond despite being sent multiple emails. Following contact, positive initial responses were received from 92 researchers, resulting in IPD being shared for 78 trials. Eighty-seven (28%) researchers declined to share data; justifications were recorded. The median time from first request to accessing data in one study was 241 (IQR 383.3) days. IPD sources included: direct from researcher, via academic trial funders repository and a website requiring remote analysis of commercial data. Where data were shared, a variety of methods were used to transfer data.

**Conclusion** Sharing of IPD from trials is desirable and a requirement of many funding bodies. However, accessing IPD faces multiple challenges including refusals to share, delays in access to data and having to perform analyses on a remote website.

**Trial registration** Not applicable.

## INTRODUCTION

The purpose of clinical research is to improve treatment and outcomes for patients, but

## Strengths and limitations of this study

- Both studies identified a large number of eligible trials across two separate disease areas, resulting in a large number of researchers being contacted.
- Trials included had academic and commercial sponsors, no restriction on country of origin and covered a wide timeframe (1985 to 2017).
- Results are likely to reflect a common experience in data sharing in vascular and dementia research.
- A limitation of this work is that only email or letter was used to contact authors due to available resources.
- It was not possible to survey non-responders to data requests, so any potential reasons for this should be considered as speculation.

also to do so ethically and efficiently, with the minimum possible number of patients subjected to experimental treatments in trials. Sharing of individual patient data (IPD) can allow multiple questions to be answered using the same data, thus preventing the unnecessary exposure of further patients to experimental treatments. There is an increasing expectation that researchers should share trial data when requested. Funding bodies such as the US National Institutes of Health and UK Medical Research Council<sup>1,2</sup> now require researchers to agree to share data before funding is granted. Further, the International Committee of Medical Journal Editors<sup>3</sup> also now require researchers to indicate their intention to share data in order for the trial paper to be published, as promoted by journals such as the *British Medical Journal* and *Public Library of Science (PLoS)* journals.<sup>4,5</sup> With respect to commercial trials, some sponsors openly share their trial data.<sup>6–8</sup> However, even with these policies in place, investigators may still face challenges in accessing IPD.<sup>9</sup> In a survey of academics who were corresponding authors for clinical trials, 88% said that they



were in favour of data sharing, with 75% believing that it should be mandatory.<sup>10</sup> However, only 18% of the researchers surveyed said their funder required them to share their data and of these, only 57% had actually shared data.<sup>10</sup>

The Stroke Trials Unit at the University of Nottingham has coordinated several studies that depend on pooled analyses of clinical or preclinical IPD: Optimising the Analysis of Acute Stroke Trials (OAST) collaboration,<sup>11–14</sup> Dipyridamole In Stroke Collaboration,<sup>15 16</sup> Community Occupational Therapy for Stroke,<sup>17</sup> NXY-059 for experimental stroke<sup>18</sup> and Progesterone for experimental stroke.<sup>19</sup> None of these projects could obtain IPD for all included trials. Two further studies are ongoing: Optimising the analysis of cognition collaboration (OA-Cog) and the Optimising the analysis of vascular prevention trials collaboration (OA-Prevention). These studies aim to identify the most efficient statistical techniques for analysing either cognition or vascular prevention data, respectively. Their analyses will be performed using IPD from randomised controlled trials involving patients with, or at risk of developing, either cognitive impairment or dementia, or vascular events such as stroke or myocardial infarction.

We report our experience of identifying and contacting chief investigators, and obtaining and analysing trial IPD for the OA-Cog and OA-Prevention studies.

## METHODS

### OA-Cog and OA-Prevention

Both studies finalised protocols and received research ethics committee waivers prior to commencing the data sharing phase. Strategies for identifying relevant trials and requesting IPD were similar and followed those used previously (OAST study).<sup>11–14</sup> First, relevant trials were identified through searches of the Cochrane Database of Systematic Reviews and PubMed using specific keywords related to the particular study (eg, OA-Cog ‘Dementia’ and ‘Alzheimer’s disease’; OA-Prevention ‘Stroke prevention’). The resulting outputs from these searches were then evaluated for relevance by PS and LJW. Detailed search strategies are not essential for this paper and will be reported as part of the individual studies.

Trials were eligible for inclusion if the primary analysis was statistically significant with a  $p$  value  $< 0.05$ , showing either benefit or harm, or were included in a systematic review showing statistically significant benefit or harm overall. Trials were excluded if they were not randomised, were neutral ( $p \geq 0.05$ ) within a neutral review, did not collect outcome data required in the studies (for OA-Cog this was global cognition score and for OA-Prevention severity of vascular event data) or met at least two of the following conditions:  $< 50$  participants; published before 1995; full publication of the primary analysis was unavailable.

## Author contact

Once a trial was deemed eligible, the lead researcher or corresponding author was contacted by email, letter or via an online application form. If an email address was not provided in the trial publication or the email address was no longer active, attempts were then made to identify current contact information. Methods included internet searches for more recent publications by the same author, or for other authors from the primary publication, or by contacting the researcher’s institution as listed in the publication. If no response was received within 4 weeks, a reminder email was sent to the same email address. Reminders were sent up to four times with a minimum of 4 weeks between each reminder.

Researchers were requested to share data on: patient demographics (eg, age, sex), medical history (eg, history of hypertension), baseline clinical data (eg, blood pressure, stroke type, functional status), outcome measures (eg, global cognition scale score, quality of life, vascular events and event dates) and trial characteristics (eg, randomisation group, length of follow-up). Once received, the IPD were combined into one master database with each trial given a unique identifier.

## Data sharing mechanisms

Multiple mechanisms exist for sharing data. IPD may be published within the main paper (only relevant if the trial has a small sample size), published directly on the web with open access, submitted by the lead researcher directly as a data file either by email or via USB/CD (typically with encryption to protect confidentiality), accessed through an academic collaboration data repository or on a website hosting both the trial data and statistical packages so that analyses have to be performed on the remote server.

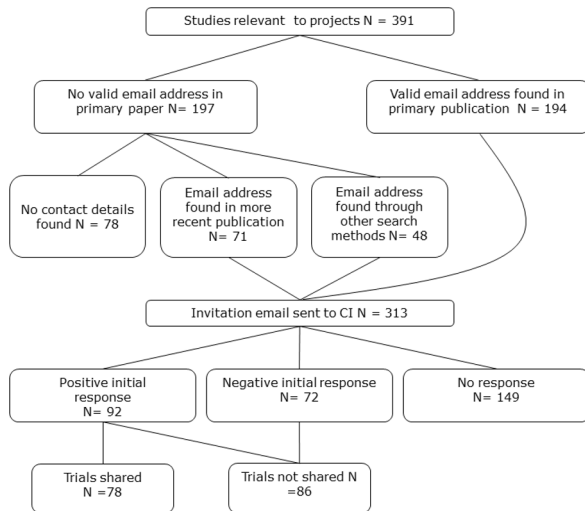
## Outcomes and statistical analyses

The main outcome for this analysis was receipt of IPD for inclusion in the OA-Cog and OA-Prevention studies. Other outcomes included response (positive or negative) and non-response rates. We compared the following trial characteristics between those that did or did not make IPD available: year of publication, sample size and sponsor type. We also collected reasons for not sharing IPD, where given. For OA-Prevention only, we collected data regarding time from initial contact date to receipt of trial data. Data are tabulated and shown as number and per cent or median with 25th and 75th centiles.

## RESULTS

### Identifying eligible trials

Of 391 trials identified as being relevant to the OA-Cog or OA-Prevention studies, an active email address for the lead researcher or a corresponding author was provided in the primary publication for 194 (50%, [figure 1](#)) and was either not given or inactive for 197 trials. Of the 197 trials, different active email addresses were identified



**Figure 1** Flow diagram—contacting authors and obtaining data. CI, Chief Investigator.

for 119 from secondary publications, other authors, the lead researcher’s institution or web searches. No email address was found for researchers of 78 trials. Invitation emails were sent to the lead researcher of the resulting 313 trials. Thirteen trials were eligible for both studies; the lead researcher was contacted using a joint email for two of these trials.

### Obtaining IPD

Of 313 invitations sent, we received positive initial responses from 92 researchers (29%), from whom 78 trial datasets were obtained. In spite of further reminders, 14 researchers did not follow through on their initial willingness to share data. One hundred and forty-nine (48%) researchers did not respond despite evidence that the email address was still in use and multiple reminders being sent. Eighty-six (27%) researchers declined to share their data (table 1); reasons included ‘operational constraints’, ‘non-availability of data’ and non-availability of the researcher (death, retirement, moved into another scientific field). Of the 13 trials who were contacted for both studies, four researchers responded positively and shared their data. However, four of the researchers did not respond at all to either of the studies and the remaining five agreed to share data with one study but not the other.

### Timings

The median time from sending the first email to successfully receiving data in the OA-Prevention study was 242 (25th, 75th centile: 21.5, 405) days. This varied between the different methods for sharing data and data source (commercial website/lead researcher/academic repository; figures 2 and 3). Time from initial contact to receiving a copy of data was shorter if sent via email or USB (median 86, 25th, 75th centile 20–405 days), but was longer if shared via an online platform (median 336, 25th, 75th centile 315–357 days) (figure 2). If the data source was a commercial website, the median time from first request to data being shared was 336 (25th, 75th centile:

**Table 1** Reasons given by researchers for not sharing data

Reason for not sharing data	N=86
Operational constraints	19
Data not available	13
Lead researcher retired/left field of study	10
Lack of communication after initial positive response	9
Company policy is to only share data from studies approved by EU and US regulatory agencies on or after 01/01/2014	7
No reason given	7
Informed consent procedures	6
Unsure whether to share data	5
Lead researcher unable to take project on/too busy to collaborate	4
Electronic databases no longer exist	2
Contract issues	2
Policy not to release IPD	1
Lack of enthusiasm for our pooling projects	1

IPD, individual patient data.

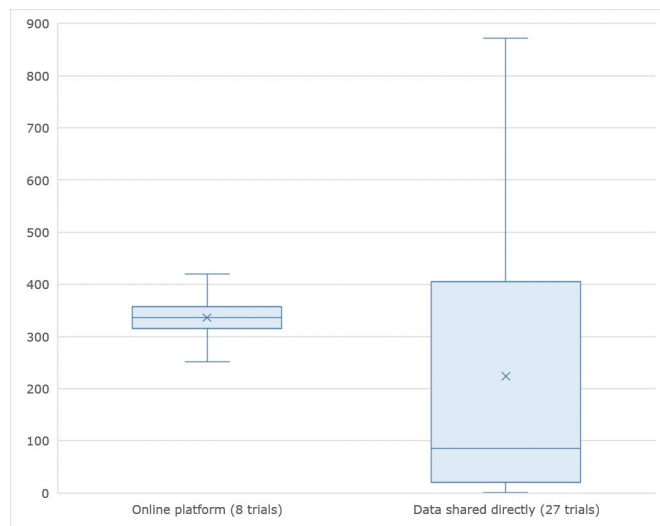
273, 336) days, for lead researcher it took 169 (25th, 75th centile: 6.5, 495.75) days and for academic repository 23 (25th, 75th centile: 20, 221.5) days (figure 3).

### Characteristics of trials received versus not received

In a comparison between positive, negative and non-responders, there was evidence of a difference in year of study publication ( $p=0.027$ ). Non-responses tended to be from newer studies (median (25th, 75th centile): 2005 (2000, 2008)) compared with positive (2003 (1999, 2007)) and negative responses (2002 (1999, 2005)). However, there was no evidence to suggest an association between year of study publication and whether IPD were eventually shared (data shared median year 2003 (1999, 2007) vs data not shared 2004 (2000, 2008),  $p=0.29$ ). Median (25th, 75th centile) sample size for trials where data were shared (969 (349, 4444)) was larger than those not shared (344 (96, 1405),  $p=0.0002$ ). However, there was no strong evidence that study size was associated with positive (973 (349, 4444)), negative (453.5 (157, 1101)) or no response (286 (68, 1439),  $p=0.061$ ) from study authors. Trials sponsored by the pharmaceutical industry were less likely to be shared compared with those sponsored by an academic institution or another source (17% vs 46%,  $p<0.0001$ ).

### Quality of data shared

The quality of the data shared with the two studies was mixed. Both studies found that on more than one occasion data were sent missing key variables (eg, randomisation allocation) that had been requested in the original application. This meant that further contact with the lead researcher was required to confirm why this data was missing and whether it was available for sharing. This

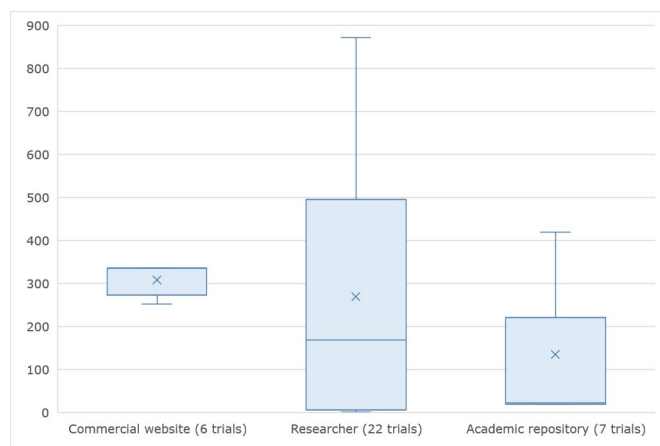


**Figure 2** Boxplot showing the time, in days, from first data request sent to receiving the data, by method of sharing.

was a time-consuming issue and also led to one trial being excluded as it could not be analysed. Occasionally, data were also sent with the variable headings and codes written in another language, without a data dictionary. This made it harder to identify the variables needed for the studies and increased the time it took to merge with the other IPD. It is also possible that this may have led to incorrect assumptions being made when converting the data into a standard format. Another approach used was to share all the trial's data in its raw format. If shared with a well-written data dictionary, this was not a problem, however, this was not always the case. One message that both projects discerned from this process is that even though inclusion of a well-written and linked data dictionary is key in data sharing, this is variably done in practice.

### Mechanisms of data sharing

The various mechanisms available for sharing data and the advantages and disadvantages associated with them are given in [table 2](#). In some cases, data are shared directly with, and analysed by, the researcher. A number



**Figure 3** Boxplot showing the time, in days, from first data request sent to receiving the data, by source.

of pharmaceutical companies have committed to sharing their data via hosted servers: Clinical Study Data Request data repository,<sup>20</sup> Yale University Open Data Access project.<sup>21</sup> In contrast to other sharing methodologies, analyses are performed on the host server with analyses having to be performed using specified software (SAS or Stata).

### IPD in main publication

IPD may be published as part of the primary publication although this is only feasible for small trials. For example, data from Graf's trial<sup>22</sup> were extracted from the publication and reformatted into an analysable dataset.

### Published data on the web

Data from completed trials may be published online so that they can be accessed free of charge, with no application to the data holder required. Data from the IST trial have been published in this way along with a data dictionary, which describes the dataset, including variable names and codes.<sup>23</sup> The trial dataset was downloaded directly from the supplementary material of the publication and was immediately ready to analyse.

### Researcher provides data on request

Researchers may store data at their institution and oversee the data sharing process each time there is a request. Typically, this required an email to the lead researcher, including details about the study. If agreed, transfer of data was made under a data sharing agreement with signature by both the data holder and the user. The researcher then provided the data in an analysable format such as an Excel spreadsheet or SPSS data file. In most cases, a description of the dataset was provided along with the data. Researchers were usually willing and able to provide answers to follow-up queries.

### Purchase of a CD of data

IPD may be on a CD/DVD that can be bought on-line. The data from the National Institute of Neurological Disorders and Stroke (NINDS) alteplase trial<sup>24</sup> was shared by this method. Once the data holder had agreed to share their data, the encrypted CD was sent by recorded post.

### Academic collaboration data repository

A number of academic collaborations exist to facilitate data sharing of randomised controlled trials in a particular disease area. The collaboration is approached via email and an on-line application form; the form requests details of the study's hypothesis, analysis methods, which trials are required, and contact details of those wishing to have access to the data. The application is reviewed by the collaboration's steering committee and, following approval, the data are provided in an analysable format. Examples include the NINDS<sup>25</sup> and NHLBI<sup>26</sup> repositories and the Alzheimer's Disease Collaborative Study (ADCS).<sup>27</sup> Some repositories include IPD without randomisation details so that on-treatment analyses may not be possible, for example, VISTA.<sup>28</sup>

**Table 2** Methods for sharing individual patient data

Method of data sharing	Example(s)	Advantages	Disadvantages	N
IPD published in main publication	GRAF <sup>22</sup>	Free. No need to contact lead researcher. No effort required by data owner once data are published.	Feasible for smaller studies only. No formal support by lead researcher.	1
IPD published on web	IST aspirin/heparin <sup>23</sup>	Free. Quick and easy to access. No application required. No effort required by data owner once data are published.	No formal support by lead researcher.	1
Lead researcher holds IPD and provides on request	Van de Winckel <sup>21</sup> PODCAST <sup>29</sup> EAF <sup>30</sup>	Free. Minimal resources required once dataset has been prepared for sharing. Lead researcher has complete control over who the data is shared with.	Requires active participation by lead researcher, that is, they need time to share data and approve each request. Assumes lead researcher can still be contacted (ie, not retired or dead). Quality of data and accompanying materials shared are variable.	46
Purchase CD of data	National Institute of Neurological Disorders and Stroke (NINDS) alteplase <sup>24</sup>	Minimal resources required once dataset has been prepared for sharing.	Cost of CD.	0
Academic collaboration data repository	Alzheimer's Disease Collaborative Study (ADCS) <sup>27</sup> NINDS <sup>25</sup> National Heart, Lung, and Blood Institute (NHLBI) <sup>26</sup>	Short application process. Responses to requests are usually quick. Application process allows more control over who has access to data. Once data has been prepared, minimal further input is required from the data owner.	NINDS, NHLBI require evidence of Research Ethics Committee approval or waiver. IPD may not be complete.	14
Commercial website	Clinical study data request.com. <sup>31</sup> Merck. <sup>32</sup> Yale University Open Data Access (YODA) Project <sup>33</sup>	Apply for data from multiple companies. Pharmaceutical companies can share data from their clinical trials via an independently monitored online system.	Long application process. Long waiting time for responses. Data needs to be analysed within the repository, this requiring upload of SAS analysis code. Often the platform only allows access to one statistical analysis software package making the data inaccessible to researchers without the necessary skills. Companies choose what data to share. Complicated, technically challenging and time-consuming process. Assumes that remote analyses will give the same results as those obtained locally.	18

IPD, individual patient data.

### Commercial website

A number of pharmaceutical companies share their data via externally hosted servers: Clinical Study Data Request data repository ([www.clinicalstudydatarequest.com](http://www.clinicalstudydatarequest.com)),<sup>20</sup> Yale University Open Data Access project (<https://yoda.yale.edu>).<sup>21</sup> Examples include the P<sub>RO</sub>FESS and JASAP trials. Access is requested via an on-line application form, this requiring information on the study's overview, rationale, aims and analysis methods, plans for dissemination of results and names of those wishing to access data. Once approved, a three-way contract between the commercial site, the company who own the data and the institution wishing to obtain the data is required. The process of negotiating the contract took several months, after which access to the remote analysis platform was granted. Some time was needed to set up access to the system and for the data user to become familiar with it. Statistical analysis code (written using SAS V.9.3 or V.9.4, or Stata V.15) was uploaded and run on the remote server. Download

of results (typically point estimates, CIs and significance values) had to be approved by the data-owning company (presumably to prevent a direct export of the IPD). Approval could take a few hours to a few days, after which the results could be exported. During the studies, one website (<https://yoda.yale.edu>) changed the host statistical software from SAS to Stata requiring statistical analysis software to be re-written. In the case of another website ([www.clinicalstudydatarequest.com](http://www.clinicalstudydatarequest.com)), two of the pharmaceutical companies that had agreed to share data with one of the studies decided to withdraw all data from the repository, causing issues with data access and required revision of the original contracts.

### DISCUSSION

The ongoing OA-Cog and OA-Prevention studies require data from completed trials to be shared. Despite difficulties in accessing data, there were researchers who



were enthusiastic and forthcoming with support for our studies and with sharing their data for 78 trials. There are now several different mechanisms for data sharing available to researchers, a number of which were utilised to share data with our studies. Unfortunately, there was a high non-response rate with 149 (48%) lead researchers not responding to an email invitation despite multiple attempts to make contact. A large number of researchers also declined to share the data from their completed trial with the studies.

The difficulties we experienced have been encountered by others conducting research projects that require IPD from completed trials to be shared. For instance, Fleetcroft *et al*<sup>9</sup> reported that of 30 eligible trials, 6 did not have available contact information for the lead researcher. Of the remaining 24, 18 (75.0%) did not respond and 5 (20.8%) declined to share their data. This could indicate that data sharing is not a high priority for many researchers. It is now expected that researchers should make available a copy of their published trial data when requested. Journals, such as the *BMJ*,<sup>4</sup> now require researchers to make their trial data available for sharing in order for the primary results to be published. In addition, some also require that authors include a data availability statement within the final publication (*PLoS*<sup>5</sup>). Additionally, funding bodies such as the US National Institutes of Health and UK Medical Research Council<sup>1 2</sup> also require researchers to agree to share their data within a given timeframe of completion of the trial before funding is granted. Such policies will help with future trials, but do not incentivise researchers of older trials to share their data.

Non-response from lead researchers was a major barrier to accessing IPD from trials. It may help that journals now require a contact email address with every publication. However, this does not solve the problem, as these studies have discovered. While the general consensus is that IPD from trials should be made available for sharing, but in reality requests for data are often missed or ignored. It is important that researchers are able to access older data. If the lead researcher moves on it should be the responsibility of the institution to continue support for potential data sharing. Ideally, it would be mandatory for the data custodian's contact details to be made available in the primary publication and possibly in the trial's registry information. Lack of resources for preparing data for sharing (table 1) is a common reason not to share IPD, as also seen in a survey of trialists.<sup>10</sup> Some methods for sharing IPD require few resources for the data owner once the initial effort of preparing the data for sharing is done, for example, when utilising commercial websites and academic trial repositories. Storing the data within a repository can also take the burden of approving a request away from the lead researcher. However, this can be costly and may not be appropriate for all trialists. Resources should be set aside to enable proper archiving and sharing of data to take place once a trial is completed. To help trialists, it would be beneficial for a universal set of standards for data preparation, storage

and sharing to be developed and promoted by organisations that mandate data sharing. Another concern many researchers have is that data from their study may not be used appropriately. It is the responsibility of the requester to provide full details regarding why the trial data are required; how it will be used and stored; how it will be published and whether the trialists will be co-authors and how and when it will be destroyed. A formal data sharing agreement/contract should be used for data security purposes. Patients may also have concerns about how their data will be used in the future. It is vital that consent from patients includes information about whether their data may be shared with other researchers and in what context. The MRC have published guidance for trialists on data sharing and informed patient consent.<sup>20</sup>

### Strengths and limitations

A strength of this research is that both studies identified a large number of eligible trials across two separate disease areas, resulting in a large number of researchers being contacted. In addition, the requested trials had both academic and commercial sponsors, no restriction on country of origin and covered a wide timeframe (1985 to 2017); therefore, our results are likely to reflect a common experience in data sharing in vascular and dementia research. Limitations include using only email or letter to contact authors due to available resources. Telephone or face to face contact methods were not utilised and might have resulted in a different outcome for some trials. Next, it was not possible to survey non-responders to data requests, so any potential reasons for this should be considered as speculation. This is unfortunate, as it means that the true barriers to data sharing are not, at this moment in time, fully known.

### CONCLUSION

In summary, there are multiple hurdles that still need to be overcome in order to facilitate more collaboration and sharing between researchers. Lack of response from researchers and unwillingness to share data make pooling studies such as these difficult to deliver in practice. Based on our experiences, data sharing still appears to depend on the enthusiasm of the lead researcher, despite it being widely agreed that IPD should be easier available. This troubling trend not only means that the results of trials cannot be checked by an independent researcher but that further patients can be subjected to experimental treatments when the data are already available. Patients have taken a risk in volunteering their time to participate in research and as researchers we owe it to them to gain as much knowledge as possible from the data they have provided. If the patient has consented to further use of their data, then the researcher has a responsibility to ensure this is done. However, whether expectations by funders and journals for investigators and authors to share data will improve this situation remains to be seen.

**Contributors** All authors are responsible for the design of the two studies. PS and LJW conducted the analyses and drafted the manuscript. PMB and AAM commented on the analyses and drafts of this paper and have seen and approved the final version.

**Funding** LW was funded in part by UK MRC ENOS (G0501797) and NIHR TARDIS (10/104/24). PS was funded in part by UK MRC ENOS (G0501797) and the Alzheimer's Society and Stroke Association PODCAST (TSA 2008/09).

**Competing interests** PMB is Stroke Association Professor of Stroke Medicine and is an NIHR Senior Investigator. PMB was Chief Investigator for the ENOS, TARDIS and PODCAST trials, which were shared with the two studies.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data sharing not applicable as no datasets generated and/or analysed for this study. No additional data available.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Lisa J Woodhouse <http://orcid.org/0000-0002-4472-1999>

#### REFERENCES

- National Institutes of Health. Final NIH statement on sharing research data, 2003. Available: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>. [Accessed 21 Jun 2017].
- Medical Research Council. Medical Research Council. MRC policy and guidance on sharing of research data from population and patient studies, 2011. Available: <https://www.mrc.ac.uk/publications/browse/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/> [Accessed 21 Jun 2017].
- Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data: a proposal from the International Committee of medical Journal editors. *JAMA* 2016;315:467–8.
- Groves T. Bmj policy on data sharing. *BMJ* 2010;340:c564.
- PLOS. Data Availability. Data sharing policy for publications in PLOS journals, 2014. Available: <http://journals.plos.org/plosone/s/data-availability> [Accessed 31 Oct 2017].
- Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. *N Engl J Med* 2013;369:475–8.
- Harrison C. Glaxosmithkline opens the door on clinical data sharing. *Nat Rev Drug Discov* 2012;11:891–2.
- Kmietowicz Z. Boehringer Ingelheim increases access to clinical trial data. *BMJ* 2014;348:g3347.
- Fleetcroft R, Ford J, Gollop ND, et al. Difficulty accessing data from randomised trials of drugs for heart failure: a call for action. *BMJ* 2015;351:h5002.
- Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among Trialists: a cross sectional survey. *BMJ* 2012;345:e7570.
- Optimising Analysis of Stroke Trials (OAST) Collaboration, Bath PMW, Gray LJ, et al. Can we improve the statistical analysis of stroke trials? statistical reanalysis of functional outcomes in stroke trials. *Stroke* 2007;38:1911–5.
- Optimising Analysis of Stroke Trials Collaboration. Calculation of sample size for stroke trials assessing functional outcome: comparison of binary and ordinal approaches. *Int J Stroke* 2008;3:78–84.
- Optimising the Analysis of Stroke Trials (OAST) Collaboration, Gray LJ, Bath PMW, et al. Should stroke trials adjust functional outcome for baseline prognostic factors? *Stroke* 2009;40:888–94.
- Bath P, Hogg C, Tracy M, et al. Calculation of numbers-needed-to-treat in parallel group trials assessing ordinal outcomes: case examples from acute stroke and stroke prevention. *Int J Stroke* 2011;6:472–9.
- Leonardi-Bee J, Bath PMW, Bousser M-G, et al. Dipyridamole for preventing recurrent ischemic stroke and other vascular events: a meta-analysis of individual patient data from randomized controlled trials. *Stroke* 2005;36:162–8.
- Halkes PHA, Gray LJ, Bath PMW, et al. Dipyridamole plus aspirin versus aspirin alone in secondary prevention after TIA or stroke: a meta-analysis by risk. *J Neurol Neurosurg Psychiatry* 2008;79:1218–23.
- Walker MF, Leonardi-Bee J, Bath P, et al. Individual patient data meta-analysis of randomized controlled trials of community occupational therapy for stroke patients. *Stroke* 2004;35:2226–32.
- Bath PMW, Gray LJ, Bath AJG, et al. Effects of NXY-059 in experimental stroke: an individual animal meta-analysis. *Br J Pharmacol* 2009;157:1157–71.
- Wong R, Renton C, Gibson CL, et al. Progesterone treatment for experimental stroke: an individual animal meta-analysis. *J Cereb Blood Flow Metab* 2013;33:1362–72.
- Tudur Smith C, Hopkins C, Sydes M, et al. Good practice principles for sharing individual participant data from publicly funded clinical trials. *Trials* 2015;16.
- Van de Winckel A, Feys H, De Weerd W, et al. Cognitive and behavioural effects of music-based exercises in patients with dementia. *Clin Rehabil* 2004;18:253–60.
- Graf A, Wallner C, Schubert V, et al. The effects of light therapy on Mini-Mental state examination scores in demented patients. *Biol Psychiatry* 2001;50:725–7.
- Sandercock PAG, Niewada M, Czlonkowska A, et al. The International stroke trial database. *Trials* 2011;12:101.
- Dachs RJ, Burton JH, Joslin J. A user's guide to the NINDS rt-PA stroke trial database. *PLoS Med* 2008;5:e113.
- National Institute for Health. NIH data sharing Repositories. A list of data repositories used by NIH, 2013. Available: [https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_repositories.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html). [Accessed 31 Oct 2017].
- National Heart Lung and Blood Institute. Biologic specimen and data Repository information coordinating center (BioLINCC). Data repository for NHLBI, 2008. Available: <https://biolincc.nhlbi.nih.gov/home/>. [Accessed 30 Jul 2014].
- National Institute on Aging (NIA) and University of California (San Diego). Alzheimer's disease Cooperative study (ADCS), 1991. Available: <https://www.adcs.org/>.
- Ali M, Bath P, Brady M, et al. Development, expansion, and use of a stroke clinical trials resource for novel exploratory analyses. *Int J Stroke* 2012;7:133–8.
- Bath PM, Scutt P, Blackburn DJ, et al. Intensive versus Guideline Blood Pressure and Lipid Lowering in Patients with Previous Stroke: Main Results from the Pilot 'Prevention of Decline in Cognition after Stroke Trial' (PODCAST) Randomised Controlled Trial. *PLoS One* 2017;12:21.
- EAFIT (European Atrial Fibrillation Trial) Study Group. Secondary prevention in non-rheumatic atrial fibrillation after transient ischaemic attack or minor stroke. *The Lancet* 1993;342:1255–62.
- Csdr. Clinical study data Request, 2013. Available: <https://www.clinicalstudydatarequest.com/>
- Merck. MERCK Clinical Trials - Policies and perspectives, 2009. Available: <https://www.merck.com/clinical-trials/policies-perspectives.html> [Accessed 31 May 2016].
- Yale University. Yale open data access (yoda) project, 2011. Available: <http://yoda.yale.edu/>