Clinical Trial Data Sharing 2019 Cross-funder Consultation: What We've Heard from Researchers

Version 1.0

Clinical Trial Data Sharing: What We've Heard from Researchers

Authors: Georgina S Humphreys[1], George Merriott[1], Rachel Knowles[2], Ben Pierson[3], Paola Quattroni[4]

[1] Wellcome Trust, London, United Kingdom

[2] Medical Research Council, London, United Kingdom

[3] Bill and Melinda Gates Foundation, United States

[4] Cancer Research UK, London, United Kingdom

Cite this as: Humphreys, Georgina; Merriott, George; Knowles, Rachel; Pierson, Ben; Quattroni, Paola (2020): Clinical Trial Data Sharing: What We've Heard from Researchers. figshare. Report.

DOI: 10.6084/m9.figshare.11847336

Contents

Executive Summary

Clinical trials are a vital part of the translational medicine pipeline that provide vast amounts of highly valuable data. When made findable, accessible, interoperable and reusable, these data can lead to significant improvements in health research and patient care. Wellcome, Cancer Research UK, the Bill and Melinda Gates Foundation and the UK Medical Research Council have partnered to understand how we can support activities that encourage data sharing.

We ran a community consultation that consisted of a survey and day-long participative workshop. The survey was completed by 174 people who work in clinical trials, including trialists, statisticians and data re-users. The workshop was attended by 33 people, working across all stages of the trial pipeline; including clinical trialists, statisticians, data managers, funders and publishers.

From the survey we identified core themes that we then picked up in the workshop. Those core themes were:

- There needs to be better guidance and support (including training to prepare and de-identify data before sharing);
- The community needs to develop and agree on standards;
- Funders should cover costs of sharing (including preparation of data, access committees, catalogues and repositories) and make it clear how to access these funds;
- Policies are important to drive change, but only half of the community supports funders mandating the sharing of all data.

Ideas that came from the workshop included ways for funders to give more structured guidance and support to researchers, ways to implement new data standards, and thoughts on credit and incentivising data sharing.

We are committed to developing and supporting activities that enable researchers to share and use clinical trial data, informed by the feedback gained during this process.

Background & Introduction

Sharing clinical trial data is a crucial activity that supports and enhances clinical research. We believe it enables a better scientific and policy response to pressing medical challenges in a variety of ways:

- Sharing allows trial results to be re-analysed; this improves reproducibility and integrity;
- It allows researchers to draw methodological insights; this improves the design of future trials;
- It allows new insights to be drawn through the meta-analyses of individual-level patient data;
- It helps policymakers ensure that they're using the best science to inform their response to health challenges;
- It prevents the duplication of trials; this saves money and prevents an unnecessary risk to participants.

Wellcome, Cancer Research UK, the Bill and Melinda Gates Foundation and the UK Medical Research Council have partnered to develop a cross-funder strategy that aims to improve levels of data sharing among our funded clinical researchers.

All four funders are currently academic members of the clinical data sharing platform ClinicalStudyDataRequest.com (CSDR). We encourage our researchers to use this platform to share their data. Wellcome also currently acts as the secretariat of the Independent Review Panel for both CSDR and Vivli (another clinical data sharing platform), ensuring a transparent and independent data access model. Despite significant investment in CSDR from pharmaceutical companies, we are currently seeing a low level of uptake from the clinical researchers that we fund.

We've therefore joined our efforts to run a series of consultative exercises to inform and facilitate our work and improve the level of data sharing by our funded clinical trialists. This report is designed to summarise the findings of these consultations.

Survey Results

Between May and June 2019, we ran an open online survey to establish researchers' views and attitudes towards clinical trial data sharing. We wanted to understand their knowledge of the resources available to them and for them to identify any challenges or barriers that we could help solve. The invitation to participate was distributed to researchers funded by Wellcome, Cancer Research UK, the Bill and Melinda Gates Foundation and the UK Medical Research Council. Researchers were also encouraged to share the survey with others. The full anonymous database of responses is available on Figshare here[1].

Participants

Among the 174 people who completed the survey:

- 60% identified as a Principal or Chief Investigator;
- 20% as a Clinical Scientist;
- 11% as a Data Manager or Data Scientist;
- The remaining 9% included PhD students, statisticians, postdocs, and technicians.

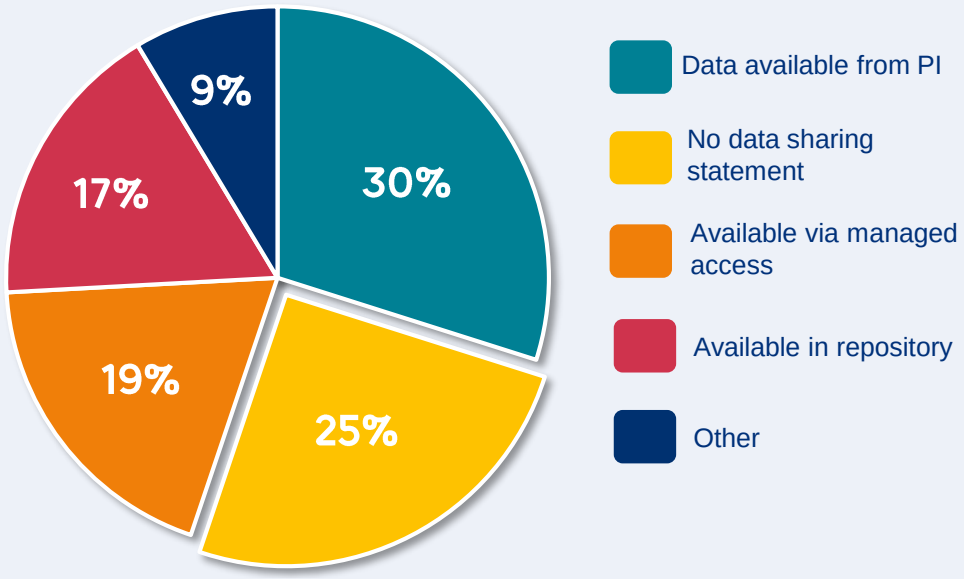
We asked introductory questions to assess the respondents' experience of running clinical trials. The majority (71%) of respondents had previously been involved in running a clinical trial, most of whom (69%) had recruited participants in the UK. We were pleased to see that there was also representation of those recruiting patients globally, including low and middle-income countries (LMICs).

Table 1:
Representation of respondents' previous/current funding
Survey, n=174

Funder	% of respondents who hold/have held a grant (n)
Medical Research Council	59.7% (104)
Wellcome Trust	54% (94)
Cancer Research UK	19% (33)
Bill and Melinda Gates Foundation	12% (21)
None of the above	17% (28)

Figure 1:
Participants' most recent data sharing statement

Survey, n=174, Question 14



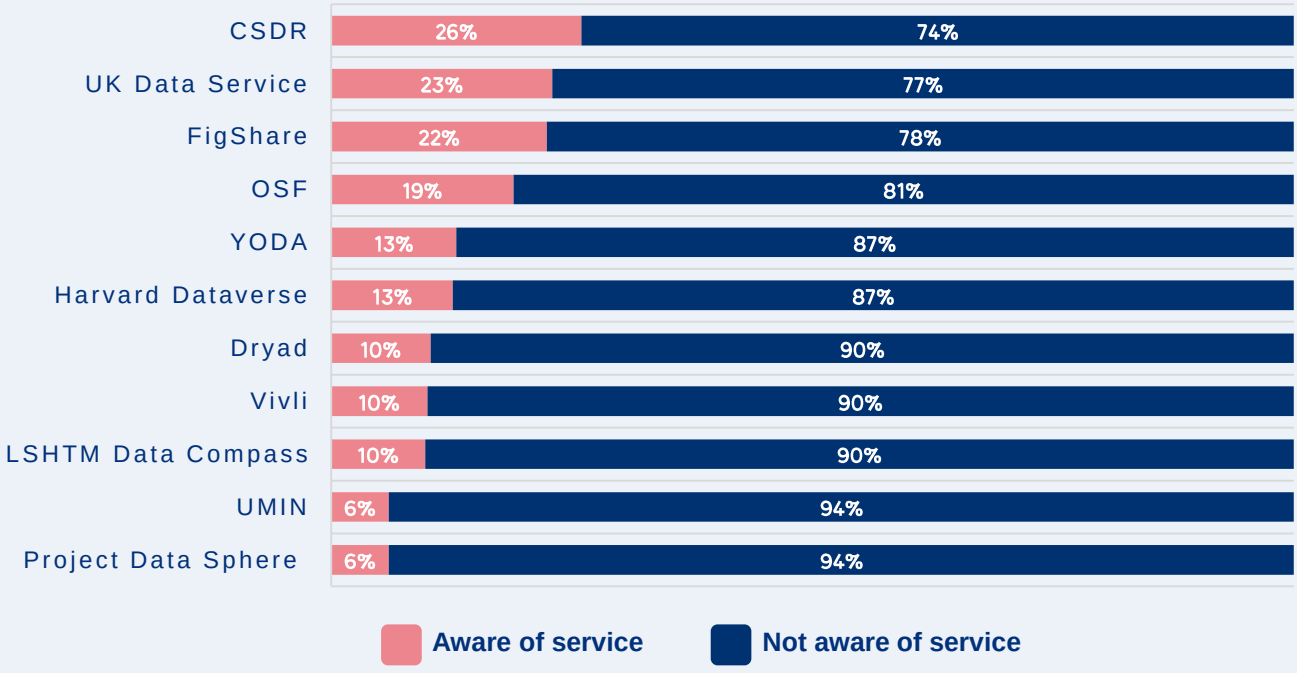
The survey was primarily distributed through our funding communities, hence 83% reported they had received funding from at least one of the four funders involved. This means however that we also captured responses from researchers beyond our grantees. The representation of funding among respondents is highlighted in Table 1.

We asked whether respondents had previously used shared data from other researchers. 61% had previously used data generated by others and 77% had shared their data in some way. Out of those who had shared data, 76% of these had shared directly with collaborators, 65% with an external team, and 53% had deposited data into a repository or listed on a catalogue.

When then asked if they included a data sharing statement in their most recent publication, 25% of respondents reported that they did not include a data sharing statement at all (Figure 1). Although we understand that not all journals currently require a data sharing statement, we expected respondents to be more engaged in data sharing and reuse than the general population of clinical researchers. Most recent data sharing statements were mixed and represent the complex landscape in clinical data sharing where various ways of sharing have emerged.

Figure 2:
Awareness of existing resources

Survey, n=174, Questions 10 and 11



Knowledge of existing resources

To assess the levels of engagement with current support services, we asked respondents whether they were aware of any of the major data sharing platforms and if they had access to data management tools or received any support for preparing data for sharing. As shown in Figure 2, the most well-known platforms for sharing data were CSDR[2], UK Data Service[3], Figshare[4] and the Open Science Framework[5].

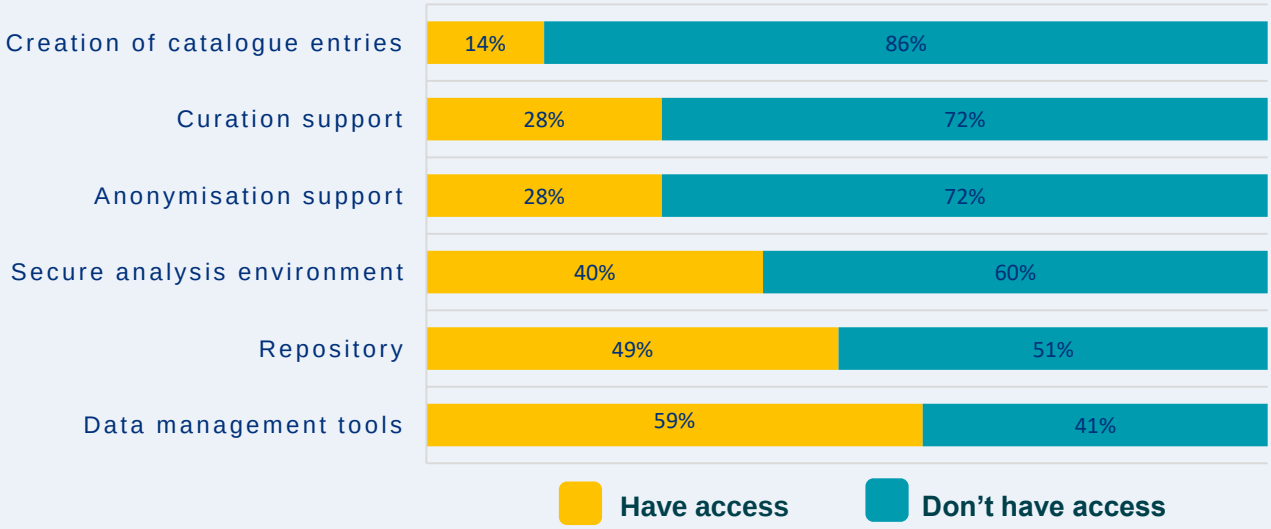
Out of the respondents who had previously run a clinical trial, 54% were unaware of any of the clinical trial specific platforms that we listed (list in Appendix I). 18% were unaware of *any* of the general platforms listed. Out of all respondents, 41% had not heard of any of the 20 listed data sharing platforms.

Respondents added 24 other platforms and repositories that they had heard of in addition to the ones listed. This clearly demonstrates the range of available options that exist for researchers, including institutional and discipline-specific repositories.

Nearly 60% of respondents said that they had access to data management tools, half had access to a repository and less than a third felt they had access to anonymisation and curation support (Figure 3). Interestingly, 13% of respondents said they did not have access to any of the resources listed, and 3% said they had access to all six resources.

Figure 3:
Current levels of access to key services

Survey, n=174, Question 2



Challenges and resource gaps

It was important to understand what the respondents viewed as being the key challenges that prevent or discourage them from sharing data. We presented a selection of nine issues, each of which were rated as 'somewhat challenging' or 'extremely challenging' by at least 60% of respondents (Figure 4). Only 13% viewed every issue as challenging, but every respondent marked at least one of the challenges as 'challenging'. The vast majority (89%) of respondents rated at least one issue as 'extremely challenging'. The top challenges highlighted were the lack of funding and the difficulty of using existing informed consent.

When asked which resources they would need or like to help them sharing data, many of the respondents felt they needed access to more support (Figure 5). This may include data curation support and access to anonymisation tools, but also simply more guidance and information about which repositories to use.

We offered respondents a list of four possible activities that funders could support to increase sharing, and the majority (84%) of respondents highlighted that funding should be provided for data management (Figure 6). 76% thought that funders should provide guidance on where and how to make data available. Interestingly, almost exactly half the respondents said funders should make data sharing a grant condition. This is in line with recent findings from the State of Open Data Survey[6], showing that more than 60% of people think that funders should make the sharing of research data part of their requirements for awarding grants. 5% of respondents ticked none of the listed options, and 22% ticked all four.

Figure 4:
Challenges facing clinical researchers

Survey, n=174, Question 4

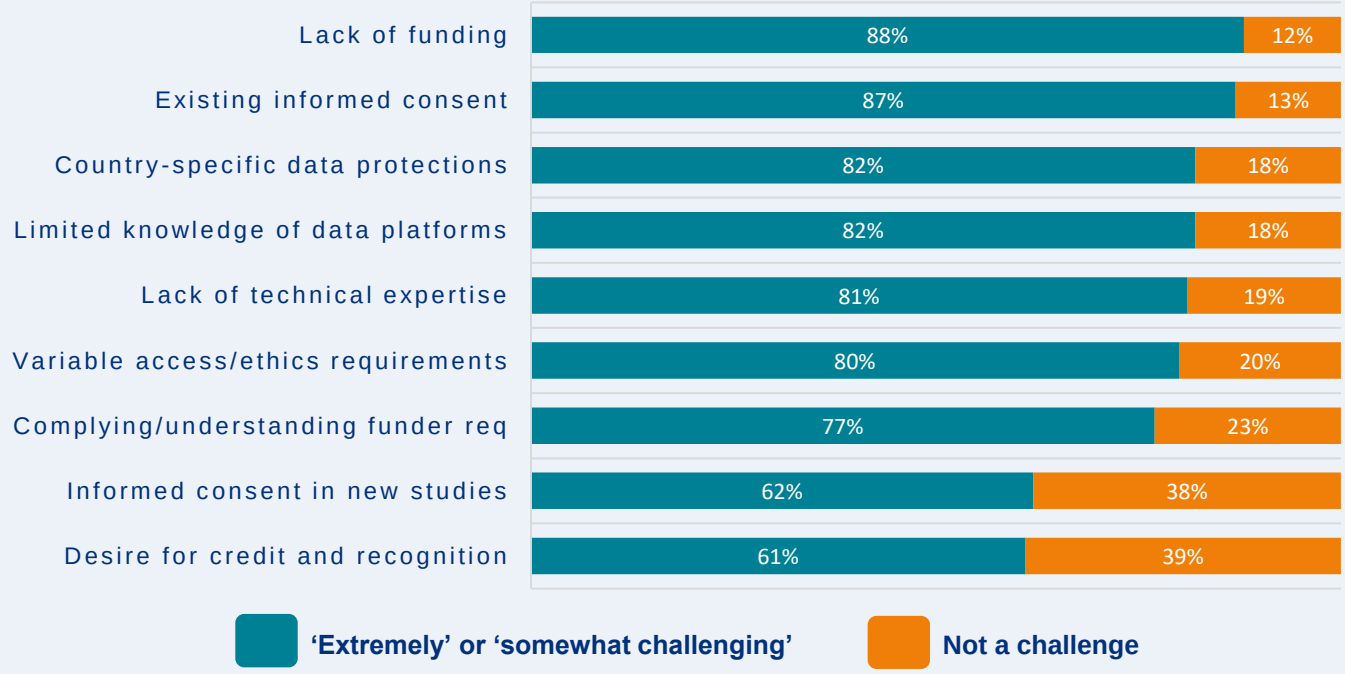


Figure 5:
Services needed for data sharing

Survey, n=174, Question 3

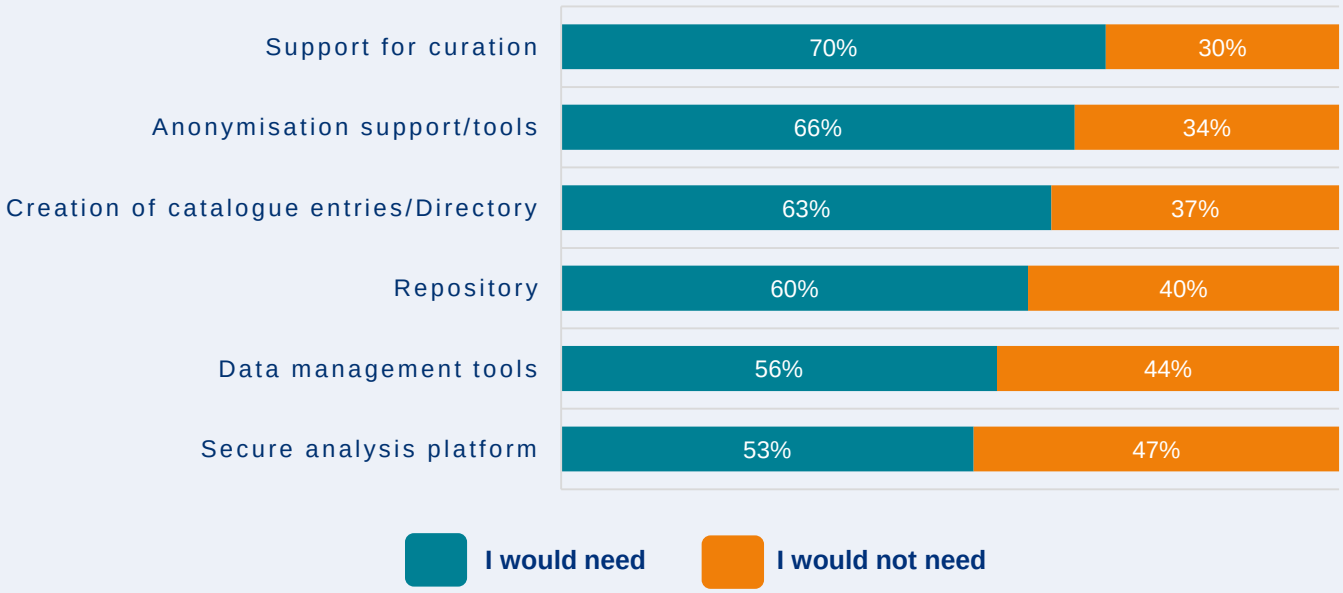
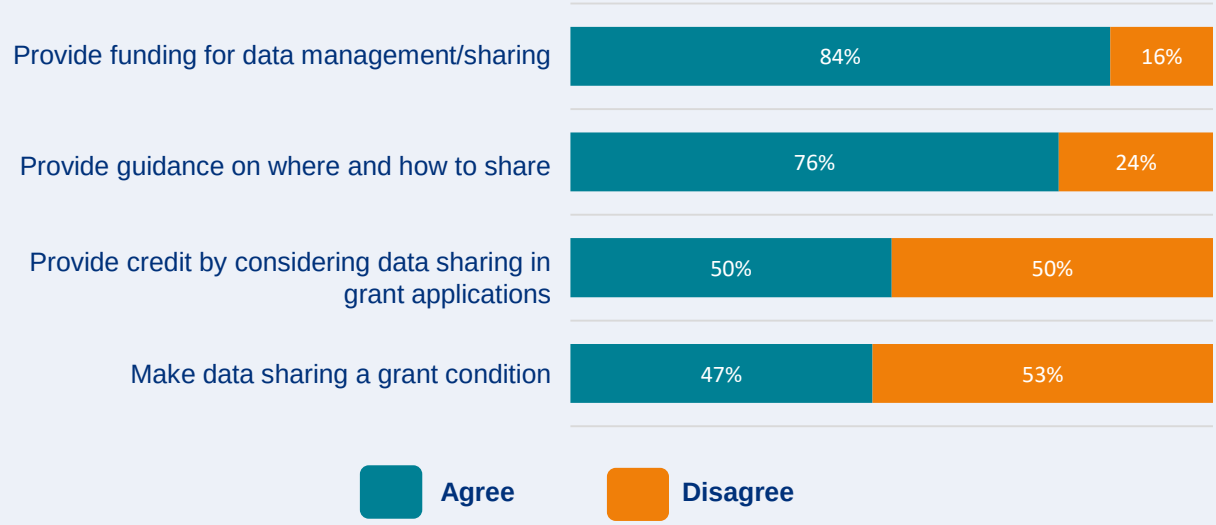


Figure 6:
What can funders do to increase sharing?
Survey, n=174, Question 12



How can funders demonstrate they value data sharing?

Finally, we allowed respondents to enter free text in response to asking how funders could demonstrate they value data sharing. We received 103 responses which broadly fell into some key themes: improving credit and incentives systems, increasing funding that supports researchers and infrastructure, mandating in grant conditions, providing robust guidance, and training.

We categorised the responses as falling into the following categories (some responses were allocated more than one category):

Category	Percentage of responses (n=103)
Funding	32%
Mandating	24%
Credit/Incentives	22%
Guidance	19%
Training	3.8%
Not categorised	21%

A selection of some of the narrative responses are included below. The full comments can be found in the published dataset.

Credit/Incentives

> *"By crediting primary data-analysis…above secondary data-analysis. The unintended consequence of data sharing is that anybody who is smart would avoid primary data collection - why pay, when you can get that for free?"*

"Have one field in grant applications: Research outputs, which should be divided into Publications, Datasets, Materials (software code etc.) and tell reviewers to value outputs equally (i.e. a publication is not necessarily worth more than a dataset)"

Funding

"Continued support for data sharing [should extend] beyond the life of the grant. Fund long term data repositories and encourage collaborative secondary data analysis."

"By providing funding. Data can only be shared if the data are well-curated, all variables are thoroughly explained in codebooks, etc. This requires staffing."

Mandating

"Carrot and stick. Enforce data sharing, but also encourage secondary use of data that acknowledges the efforts of the primary collector"

Guidance

"Putting emphasis on guidance, in my case ignorance is the main obstacle, not lack of willingness"

Training

"Promote a positive cultural change, with training /workshops to allow also scientists from LMICs [low and middle-income countries] to be able to benefit from this cultural change. Avoid the segregation between those collecting data and those analysing data."

Workshop

On the 17th October 2019 we held a one-day workshop in London (UK) to discuss advances and opportunities in the clinical trial data sharing landscape. The workshop focussed on what future activities the funders could support to drive the outputs of funded research to be open and accessible, to have the greatest possible impact. Participants included a range of roles that work on trials and/or secondary analysis of data from others, as well as researcher resource services, and publishers (the full attendee list can be found in Appendix II). The day featured invited presentations and discussions to prioritise activities around four themes:

1. Funding services
2. Development of models for credit/incentives to share
3. Guidance
4. Policies

The final agenda for the day can be found in Appendix III.

We had some lively discussion and a range of experience and expertise in the room to inform our priorities.

General themes discussed

Despite the recent emphasis on clinical trials transparency and the growing initiatives aimed at facilitating sharing of clinical trial data[7,8,9,10,11], participants highlighted that several challenges still exist when it comes to sharing data. This was consistent with the survey results. Specifically, pressure on resources and personnel, workload concerns (especially if sharing legacy data is required), and responsibility placed on statisticians to manage the risk of re-identification were among the main concerns raised at the workshop.

1. Support, guidance and tools

Many workshop participants highlighted the need of new support, guidance and tools to enable effective data sharing.

The risk of re-identification was raised as a key issue and it was felt that more guidance is needed to help researchers feel confident in sharing sensitive data. Training is also needed to prepare and anonymise datasets while ensuring that data can be re-used. To partly solve these issues, the UK Data Service[12] workshops and tools, and the material and training provided by the Global Health Network[13] were highlighted as valuable resources.

Consent is an issue that was mentioned in the survey and a concern that was echoed in the workshop. Specifically, the challenge exists in situations where consent for reuse of data has not been clearly and consistently recorded for a trial dataset. Standardisation of informed consent procedures and clear advice about the consent wording to be used in all trials so that data can be shared would ensure data are always available for reuse.

The importance of using consistent standards to enable interoperability and data re-use was also discussed, and the Clinical Data Interchange Standards Consortium (CDISC) was mentioned as a possible standards provider, although the complexity of this system was also acknowledged.

Participants thoroughly discussed anonymisation techniques and the cost of anonymisation. Catrin Tudur-Smith and colleagues[14] recently estimated that preparing data for sharing from clinical trials can take 40-50 hours of personnel time with a total associated cost of approximately £3,000. However, the cost of data preparation may be higher if the original dataset is particularly complex, old, or has not used recognised data standards for collection and archiving. As previously mentioned, in the survey feedback we saw that most respondents perceived costs as a significant barrier to sharing their data.

2. Incentivisation and providing credit

Another main topic discussed during the day was the importance of providing incentives for researchers to share their data.

Currently, the costs and efforts to prepare and share data with good metadata and the risk of re-identification liability present significant drawbacks to data sharing. We heard that the incentives for researchers to share data must encourage them to share, despite these drawbacks.

Often researchers are worried about not receiving appropriate acknowledgment and academic reward when other researchers use their data and publish papers. Data citation, co-authorship, active collaboration and emphasis of the contribution of data managers and statisticians on data papers could be part of the solution.

Liz Allen and Jonathan Threlfall (F1000) presented the CRediT contribution role taxonomy[15] at the workshop, a high-level taxonomy adopted now by more than 30 publishers, including Elsevier, PLOS, BMJ and Wiley. The taxonomy includes 14 roles to represent contributors to scientific scholarly outputs. The idea is to bring visibility and recognition of the different contributions of researchers, across all aspects of the research being reported, including labels such as ‘data curation’ and ‘statistical analysis’.

Finally, we heard that showcasing and promoting case studies to demonstrate the value of sharing and re-using data is also a good way to recognise the efforts of those who have shared more widely, beyond their direct collaborators. This can illustrate the potential to advance scientific progress by increasing the potential for data re-use.

Workshop pitches

To wrap up discussions into some tangible outcomes, we asked the participants to split into groups and work on key ideas that they felt would increase clinical data sharing. The five groups came up with the following ideas:

Pitch 1 – Improving findability:

- Data must be findable by people or machines, therefore developing standardised metadata catalogues are important;
- Consent for sharing data needs to be included in the design of future trials, using wording that is future-proofed and works well across domains;
- A data management plan should be required at the start of every trial and be updated throughout the course of the trial. This should include how to make data discoverable, such as by placing assets in repositories and considering the access model being used;
- Help research offices and institutions understand why it's important to share data, to change the culture and institutional policies. Funders could encourage research organisations (who are the trial sponsors) to effect change.

Pitch 2 – Introduction of data sharing support roles:

- A 'Data FAIRy' should be embedded in every organisation to provide advice and support for data sharing, such as advice on DOIs, standards, quality and technical assistance;
- Funders should support the development of case studies to promote data sharing, pilot data publishing, and train individuals.

Pitch 3 – Synthetic databases:

- The balance of work is currently on the data contributor, so the barrier is quite high, especially for legacy data;
- Could avoid handing over data by creating analysis code using synthetic datasets, then applying this to real data;
- Requestors could share code with data contributors who then run the analysis on real data which avoids re-identification risk;
- This might not be suitable for all data and statistical models may not work on synthetic data. Users may need training too.

Pitch 4 – Changes to training and accreditation:

- Data curation needs accreditation systems like Good Clinical Practice, so it meets a minimum standard;
- It would be useful to have international agreement on broad consent statements;
- We need to agree the level of acceptable risk and provide guidance in terms of anonymisation standards.

Pitch 5 – Development of new standards:

- To develop and adopt a set of clinical data standards. For example, CDISC could be simplified for the academic research community. This would require input from all stakeholders, including funders, academics and institutions.

Key messages from the survey and workshop

Both the online survey and the workshop highlighted some key areas that remain challenging for researchers who are trying to share their clinical trial data. There were consistent themes across many of the responses and comments we heard. This included:

- The need for guidance and support on where and how to share data;
- The need for increased provision of training and resources to guide anonymisation;
- The need for development of approaches to (1) determine and (2) mitigate the risks for each dataset prior to sharing;
- The agreement and adoption of data standards to improve interoperability;
- The development of more evidence on the impact of data sharing and documenting case studies to highlight the value of sharing;
- Building systems to incentivise and reward sharers is critical. Funders and research organisations should take a leadership role;
- Policies play a role in driving behaviour change, but mandating data sharing is only supported by half of the respondents.

Takeaways & Next Steps

This process produced several valuable insights that we will use to inform our future work in encouraging clinical trial data sharing. It also provided some evidence of the current problems encountered by researchers in sharing data, that will help us prioritise our efforts.

Both the workshop and the survey highlighted the lack of funding as one of the main challenges to data sharing. This was surprising, as all of the participating funders already allow applicants to include data sharing costs within schemes, including clinical trials. Another unexpected finding was the low levels of awareness of broad-based repositories as well as clinical trial-specific repositories.

As a group of funders, we are committed to developing future activities to improve the level of clinical trial data sharing. One of the next steps might be to review how grantees make use of the availability of existing funding and determine how we can increase our support for researchers to access this funding when they make future grant applications. Our ambition is that no researcher cites lack of funding as a barrier that prevents them from sharing their data. We also commit to providing the appropriate guidance and support to help researchers achieve data sharing objectives. Many of the challenges and issues with data sharing could be resolved with better prospective planning for data outputs as part of the trial design and granting process. We will therefore collectively assess the other types of guidance about data management, anonymisation and sharing that we provide to grantees and make improvements based on the insights drawn from this consultation. We will also review ways to promote the repositories available to researchers by providing clear and helpful guidance at the application stage.

Currently, all four funders are members of the data sharing platform CSDR and Wellcome continues to provide an Independent Review Panel for both CSDR and Vivli. We'll continue to encourage our grantees to use these services but also allow them to select and use other services as they see appropriate.

Finally, we remain committed to ensuring that our funded researchers (both clinical and non-clinical) publish data in accordance with the FAIR principles. While all our organisations have already implemented data-sharing policies and made data sharing part of their grant terms and conditions, we note there is low awareness of these policies and data sharing is not yet the norm. We'll continue to ensure that our policies emphasise these principles, and are committed to incentivise open behaviour among our researchers. Signing the San Francisco Declaration on Research Assessment (DORA)[16] and recognising that datasets are primary research outputs as valuable as journal publications are the first steps in this direction.

Acknowledgements

We would like to thank those that took the time to participate in this process by responding to the online survey and participating at the workshop (workshop attendees listed in Appendix II). We really appreciate your considered responses and will continue to use the feedback we've received to inform our future efforts.

References

1. Humphreys, Georgina; Merriott, George; Knowles, Rachel; Pierson, Ben; Quattroni, Paola (2020): 2019 Clinical Trial Data Sharing Survey Results - Data. figshare. Dataset. https://doi.org/10.6084/m9.figshare.11603295.v1
2. https://www.clinicalstudydatarequest.com
3. https://www.ukdataservice.ac.uk
4. https://Figshare.com
5. https:/osf.io
6. Science, Digital; Fane, Briony; Ayris, Paul; Hahnel, Mark; Hrynaszkiewicz, Iain; Baynes, Grace; et al. (2019): The State of Open Data Report 2019. figshare. Report. https://doi.org/10.6084/m9.figshare.9980783.v2
7. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/1480/1480.pdf
8. https://www.ncbi.nlm.nih.gov/books/NBK137816/
9. https://yoda.yale.edu/
10. https://www.clinicalstudydatarequest.com
11. https://vivli.org
12. https://ukdataservice.ac.uk/
13. https://tghn.org/
14. Tudur Smith, C., Nevitt, S., Appelbe, D. et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. Trials 18, 319 (2017) doi:10.1186/s13063-017-2067-4
15. https://casrai.org/credit/
16. https://sfdora.org/

Appendix I. Survey questions 10 & 11

These questions were designed to understand the level of awareness of current repository services and data sharing platforms available to clinical researchers.

The clinical-trial specific platforms we listed were:

- CSDR
- Project Data Sphere
- YODA
- UMIN
- Vivli

The general platforms we listed were:

- ArrayExpress
- DRUM
- EASY
- Figshare
- CPSR
- University of Bath Repository
- SND
- B2Share
- Dryad
- Edinburgh Datashare
- Harvard Dataverse
- LSHTM Data Compass
- OSF
- UK Data Service
- Zenodo

Appendix ll. Attendee list from workshop on 17th November 2019

Funders
Ben Pierson – Bill and Melinda Gates Foundation
Dottie Goble – NIHR
George Merriott – Wellcome
Georgie Humphreys – Wellcome
Hanna McEvoy – Cancer Research UK
James Donaldson-Briggs – Cancer Research UK
Paola Quattroni – Cancer Research UK
Rachel Knowles – Medical Research Council
Rebecca Haines (Dial-in) – Cancer Research UK
Tim Kinkead – Consultant for Bill and Melinda Gates Foundation

Speakers
Alex Bailey – MRC Regulatory Support Centre
Catrin Tudur-Smith – University of Liverpool
Jonathan Threlfall – F1000
Liz Allen (Dial-in) – F1000
Louise Corti - UKDS
Phaik Yeong Cheah – MORU
Philippe Rocca-Serra – Fairsharing.org/University of Oxford

Attendees
Belul Shifa – Guy's and St Thomas' NHS Foundation Trust
Chris Matthews – King's College London
Claire Lawless – University of Glasgow
David McAllister – University of Glasgow
Gill Booth – University of Leeds
Helena Wilcox – TGHN
Kalynn Kennon – IDDO
Lilian Mwango (Dial-in) - KEMRI
Mari Crotty – Guy's and St Thomas' NHS Foundation Trust
Marianne Munene (Dial-in) – KEMRI
Mark Kelson – University of Exeter
Nicky Gower – UCL
Rebecca Gallagher – Belfast ECMC
Sam Driver – TGHN
Sheela Medahunsi – Guy's and St Thomas' NHS Foundation Trust
Susannah Condie – Southampton

Appendix III. Agenda from Workshop on 17th October 2019

Our Policies and Why It's Worth It
Paola Quattroni/Rachel Knowles
CRUK/MRC

What the survey responses told us
George Merriott
Wellcome Trust

Session 1: Researcher Experience
Data Curation and Preserving Legacy Data for Sharing and Analysis
Prof. Catrin Tudur-Smith
University of Liverpool

Discussion – challenges/solutions:
- What are the challenges in preparing/curating data?
- What tools have you found useful?
- Do we need agreed international standards?
- How can funders promote data use?

Session 2: Services and Support
How to share clinical data aligned to the FAIR principles
Dr Phillipe Rocca-Serra
FAIRsharing.org

Tools for data curation and anonymisation
Louise Corti
UK Data Service and the University of Essex

Anonymisation and historic content issues
Dr Alex Bailey
Medical Research Council, Regulatory Support Centre

Institutional Data Access Committees
Dr Phaik Yeong Cheah
University of Oxford

Discussion:
- Which are the most powerful tools and what are we missing?
- FAIR standards for repositories/datasets
- Data access models e.g. DACs, IRPs

Session 3: Incentives/Credit
Crediting researchers and the CRediT taxonomy
Dr Liz Allen and Jonathan Threlfall
F1000

Discussion – how might we:
- Support researchers with appropriate credit mechanisms?
- Better promote crediting mechanisms?

Ideas Workshop:
Based on the day's discussions, develop a new solution that would promote and support clinical trial data sharing.