British Heart Foundation Data Science Centre

Led by Health Data Research UK

Report – How to facilitate the use of healthcare systems data in cardiovascular clinical trials

MARCH 2022

British Heart Foundation Data Science Centre
Led by Health Data Research UK

Report – How to facilitate use of healthcare systems data in cardiovascular clinical trials

Table of Contents

CONTENTS

1 Executive summary

The use of healthcare systems data could facilitate cheaper and more efficient clinical trials while maintaining the highest standards. We report the BHF Data Science Centre's findings which highlight barriers to the use of healthcare systems data in cardiovascular (CVD) clinical trials. These include: difficulties discovering what datasets are available and how to access these; navigating the complex bureaucracy, consent, and ethics, and obtaining information or guidance on this; and challenges associated with the costs of accessing data, understanding the costing model, and obtaining transparent and consistent cost estimates. We outline our plans to address the challenges and facilitate the use of healthcare systems data in CVD clinical trials, ultimately increasing the proportion of trials that use such data.

2 Background

Clinical trials are pivotal to evidence-based healthcare improvements, through the testing of drugs, vaccines, medical devices, surgical and others interventional procedures, diagnostic tests and a range of complex interventions. Unfortunately, the development of new interventions is extremely expensive, with clinical trials being by far the most expensive part[1,2]. A substantial part of this cost is involved in trials gathering their own data, despite the UK's rich and population-wide healthcare systems data, collected routinely as part of healthcare. The use of healthcare systems data could improve the efficiency and cost of randomised clinical trials. However, there are substantial challenges, with a very small proportion (~3%) of contemporary clinical trials estimated to be using this data[3].

A major initiative led by the British Heart Foundation (BHF) Data Science Centre, in coordination with Health Data Research UK and NHS DigiTrials, aims to streamline data-enabled cardiovascular (CVD) clinical trials. To explore the use of healthcare systems data in cardiovascular clinical trials we initially carried out a survey[4], which highlighted barriers to the use of healthcare systems data in CVD clinical trials. To delve deeper into these challenges and explore possible solutions, we organised a workshop in collaboration with the National Cardiac Surgery Clinical Trials Initiative[5]. Insights from these activities, along with wider discussions, are being used to shape and prioritise the BHF Data Science Centre's work in this area, with the aim of increasing the number and proportion of cardiovascular clinical trials utilising healthcare systems data.

3 Workshop – using health data in cardiac surgery clinical trials

3.1 Workshop aims

- To better understand the challenges faced in the use of linked health data in cardiac surgery trials
- To inform cardiac surgery trials teams of the role of the BHF Data Science Centre
- To inform how the BHF Data Science Centre can evolve its plans to best support cardiac surgery trials and the wider clinical trials community
- To work towards shared solutions

3.2 Workshop organisation

The two-part workshop was held virtually on the mornings of the 8th and 15th October 2021. Along with members of the BHF Data Science Centre (hereafter referred to as BHF DSC) team and the National Cardiac Surgery Clinical Trials Initiative operational executive team, the workshop brought together cardiac surgery trials team representatives, including leads and members involved in trial planning, management, coordination, operations, data management and analysis. In addition, four patient and public representatives from the National Cardiac Surgery Clinical Trials Initiative attended the workshop to provide input from a patient and public perspective, particularly to the development

of solutions, and to feedback to the wider group of patient and public representatives within the Initiative. A list of workshop participants is included in Appendix A (Section 8, Additional material).

The first part of the workshop focused on defining the challenges faced in accessing and using healthcare systems data in cardiac surgery trials. The second part focused on developing shared solutions to the challenges identified. The workshop agenda is included in Appendix B (Section 8, Additional material).

3.3 Part 1 - Defining challenges

The first part of the workshop opened with a "Setting the scene" session, which opened with introductions to the BHF DSC, the use of healthcare systems data in clinical trials and the challenges identified in our survey. This was followed by an introduction to the National Cardiac Surgery Clinical Trials Initiative and lightning talks from cardiac surgery trials teams to share information about currently planned or ongoing trials that are using or plan to use healthcare systems data.

The remainder consisted of an interactive breakout session to gather input on the challenges faced in accessing and using healthcare systems data in cardiac surgery trials. During the breakout session, workshop participants were split into four groups, each with a chair and additional facilitator. Participants were prompted to explore in more detail the main challenges identified in the aforementioned survey we conducted earlier this year,[4] challenges associated with gaining approval/access to healthcare systems data, challenges related to the cost of accessing data and the value associated with that data and challenges associated with data processing and analysing. Participants were also encouraged to raise additional issues. An interactive whiteboard tool, Mural, was used to gather participant input. Screenshots of the Mural boards from each group are included in Appendix C (Section 8, Additional material).

3.4 Challenges identified

Approximately 300 statements were gathered on the challenges faced in accessing and using healthcare systems data in cardiac surgery trials across the four breakout groups. A list of all statements gathered during the breakout groups is available[6]. The majority of the challenges identified in the workshop can be grouped into three key areas:

- Challenges associated with gaining access to healthcare systems data for use in clinical trials, navigating the bureaucracy, consent and ethics, and obtaining information or guidance on this
- Challenges relating to the cost of accessing data and the value associated with that data
- Challenges relating to using the data, including defining the data required, data availability and quality, and data processing and analysis

Additional challenges which fall outside of these areas included linkage across datasets, making the case for the use of healthcare systems data in funding applications, and ensuring inclusion of the public/patient voice and perception on the use of healthcare systems data in clinical trials. A key theme running across these areas related to a lack of available information, understanding, transparency and best practice. Discussions also identified a need for evidence to support the ability of healthcare systems data to support clinical trials.

Statements were manually reviewed and grouped into "challenge themes" according to the likely cause of the challenges stated. For example, the statements "Need clarity of info on which datasets are available from which provider" and "Knowing what datasets are available, what is available in each dataset, how to access them, can they be linked" were grouped under "Not clear what datasets are available and from which providers". The 25 "challenge themes" with the most associated statements are shown in Table 1, including example statements that were grouped under each theme.

Table 1. Challenge themes identified

Each challenge theme with the number of statements grouped under each challenge theme and example statements. The challenge themes selected for the development of shared solutions are shaded in grey with italic font.

Challenge theme	Number of statements	Example of statements grouped under theme
Costs are opaque and vary hugely	*19*	*Why the costs of accessing the data are so high (and variable!)*
		Costs vary wildly from one application to the next. (10 fold differences at times £5k to £50k. At times felt it was being run for profit rather than research support.
		Rationale for costs that are provided would be really helpful, especially from NHSD but also other data custodians
Linkage	12	Need to understand linkage of datasets
		Is it possible to accurately link data across multiple datasets in a standardised fashion?
		Difficult to link datasets due to need for patient anonymity.
Takes too long to get data, or unclear how long it will take	12	Everything takes too long!
		Lack of confidence we would get data in time for things like DMEC committees
		Long term follow-up for a trial has taken two years from NHS Digital - no specific reason given beside pressure of work
Patient/public voice	11	How do research teams include the patient voice in data science?
		Do organisations need to make patients/carers aware of this use for research purposes?
		How do the patient population feel about use of routinely collected data?
Bureaucracy, consent, ethics	*10*	*Processes - differ between NHS organisations - different requirements*
		Difficult / inflexible requirements for NHS Digital
		What is the right form of wording for consent? Is this consistent across providers?
Incomplete/inaccurate/quality of data	9	Accuracy and completeness of data
		Quality of data may differ between organisations
		Limitations of having to assume that a lack of code means that a patient does not have that diagnosis or treatment.

Challenge theme	Number of statements	Example of statements grouped under theme
Learning from trials that have successfully used datasets	*9*	*What are the short-cuts!!!???*
		Some trials groups seem to access this data all the time - what's different about them
		Real life examples case studies of successful use of digital data and how issues have been resolved
Deriving outcomes from routinely collected data/codelists	*8*	*Identifying patient phenotypes from routine datasets*
		Need to standardise the definition of outcomes and characteristics used in cardiac surgery trials, e.g. multiple codes used for diabetes, there should be published consensus on what codes should be used going forward
		Exactly what are the codelists for key outcome measures
Not clear what datasets are available and from which providers	*8*	*Need clarity of info on which datasets are available from which provider*
		Knowing what datasets are available, what is available in each dataset, how to access them, can they be linked
		We should have a good catalogue of what data are robustly available
Making the case for routinely-collected data to funders/on grant applications	*8*	*Grants committees are not convinced that routinely collected data is adequate for clinical trials*
		How you sell using RCHD to the funders
		How to describe digital data access for a successful grant application
Skills/expertise required to process data	*8*	*How to find an experienced data scientist/ statistician to deliver the work*
		What training courses are available
		Expertise for processing different datasets (e.g. HES, CPRD)
Challenges with data cleaning, formatting and integrating	6	Data is rarely analysis ready and integrating different datasets, formatted differently, is very difficult
		Data cleaning takes more time
		Electronic health data is complicated to clean and analyse due to the size and longitudinal data
Difficult to get help/guidance	6	Data providers have limited understanding of clinical trials and their issues
		Clear information on process of how to access datasets
		There needs to be clear guidance for people applying for the first time, what to expect, how to mitigate delays etc.

Challenge theme	Number of statements	Example of statements grouped under theme
Difficult to select/define the data you require (data dictionary not accurate/comprehensive)	6	Defining the data required is challenging without a clear and simple description of the sources available and what they contain...easy to use overviews as well as detailed data dictionaries would probably help
		Discerning the fields needed is challenging if you are relatively new to the data
		Differences on data definition
Application process is too complex	6	DARs application has several stages before HES data can be accessed but difficult to know what these are prior to submission
		The level of information required in DARs application is highly specific however this is not explained again prior to submission. Guidance only provided to us for a trial after an initial rejection. It seems difficult to complete this application despite the participants already giving consent for their data to be used.
		Requirement to apply for data access for each dataset

Eight "challenge themes" were selected for the development of shared solutions during part 2 of the workshop based on the following criteria:

- Themes that were frequently raised as creating a challenge in the survey, during the workshop, in the literature and through the experience of the BHF DSC team
- Challenges that present blockers near the beginning of the process e.g. gaining access to healthcare systems data
- Ability of the BHF DSC to address the challenge
- Perceived positive impact that addressing the challenge would have on the ability of trials teams to access and use healthcare systems data in clinical trials

We translated each of these "challenge themes" into a "How might we….?" question that captures what would be required to overcome the challenge. Table 2 includes the eight "challenge themes" selected and the "How might we…?" question for each. We also added an additional question "How might the BHF DSC best serve the community in this area of data-enabled trials?" to obtain community input on the role of the BHF DSC in this area.

Challenge theme	"How might we….?" Question	Mural board (App D)
Bureaucracy, consent, ethics	How might we make the necessary bureaucracy, obtaining consent, and ethics easier to navigate?	D1
Costs are not transparent and vary hugely	How might we make it easier to obtain an accurate and consistent cost estimate?	D2
Deriving outcomes from healthcare systems data/codelists	How might we make it easier to consistently derive outcomes from routinely-collected health data?	D3
Making the case for healthcare systems data to funders/on grant applications	How might we make it easier make the case for routinely-collected health data to funders/on grant applications?	D4
Not clear what datasets are available and from which providers	How might we make clear what datasets are available and from what providers?	D5, D6, D7, D8
Skills/expertise required to process data	How might we ensure that the staff working on clinical trials have the skills necessary to use routinely-collected health data?	D9
Difficult to get help/guidance	How might we ensure that clear guidance is available to help trials teams navigate the process of gaining access to RCHD?	D10
Learning from trials that have successfully used datasets	How might we provide evidence that routinely-collected health data can successfully support clinical trials?	D11
	How might the BHF DSC best serve the community in this area of data-enabled trials?	D12, D13

Table 2. Challenge themes selected for development of solutions.
Each challenge is shown together with the "How might we…?" question used to explore each challenge theme during the interactive breakout sessions in part 2 of the workshop. The additional question "How

might the BHF DSC best serve the community in this area of data-enabled trials?" was added to obtain community input on the role of the BHF DSC in this area.

3.5 Part 2 - Working towards shared solutions

Part 2 of the workshop opened with an interactive breakout session to develop shared solutions to the challenges selected from part 1. We used the interactive whiteboard tool, Mural, to capture the discussion. Screenshots of the Mural boards from each group are included in Appendix D (Section 8, Additional material). All statements added to the Mural boards during this session are included in Appendix E (Section 8, Additional Material).

The nine "How might we…?" questions were divided between the four groups, so that each question was explored by at least one group. During the breakout session, participants first discussed *why* it was important to address the challenge. Considering the reasons why resolving the challenge is important ensured that the end goal was kept in sight and that solutions were focused on achieving this. Participants then considered *how* we might address the challenge, while keeping the end goal of *why* we are trying to achieve this in mind, exploring different options and delving deep into each proposed solution.

4 Discussion/Conclusions

The challenges associated with the use of healthcare systems data in clinical trials are well documented, including by HDR UK and collaborators[7].

We believe that the input gathered in the workshop can be regarded as representative of the general challenges to the access and use of healthcare systems data faced by cardiovascular trials teams. While the workshop was centred around members of the National Cardiac Surgery Clinical Trials Initiative, members involved in all aspects of clinical trials participated in the workshop. This included cardiologists, cardiac surgeons, trial managers, statisticians, database managers and methodologists. In addition, members were encouraged to invite any colleagues involved in applying for access to, processing or analysing healthcare systems data for use in cardiovascular clinical trials. We also note that the experience of many participants will extend beyond cardiac surgery trials, to trials investigating the entire circulatory system and other diseases.

The challenges identified in this workshop align closely with those identified through our survey, in wider discussions with trials teams and published reports [3]. Indeed, none of the challenges raised were considered unique or specific to cardiac surgery trials. The general homogeneity of the challenges identified reflects the common requirements, processes and challenges for data access, costs, and datasets, regardless of disease area.

The majority of the challenges we have identified occur during the early stages of the process, such as gaining access to healthcare systems data and the costs involved with this. These challenges may be effectively blocking the use of healthcare systems data in clinical trials, with many trials currently "dropping out" of the process at this point or not attempting to use healthcare systems data due to perceived challenges. It is only when challenges occurring early in the process are resolved, and more clinical trials are accessing and using healthcare systems data, that we will obtain a more accurate representation of all challenges involved. It is imperative that issues preventing clinical trials from accessing healthcare systems data are resolved to enable an accurate assessment of the utility of these data to support outcome ascertainment in clinical trials.

The exploration of the challenges identified and development of solutions during part 2 of the workshop, generated rich and diverse solutions, while also capturing some specific user requirements for the proposed solutions. It also revealed solutions that could make notable improvements for users

across multiple challenges. For example, creating a central resource that makes available information on healthcare systems datasets, including how to access them, what variables each dataset includes, how these variables can be combined to derive outcome measures, and links to any available training resources, would address the challenges "Not clear what datasets are available and from which providers", "Difficult to get help/guidance" and "Deriving outcomes from healthcare systems data/codelists". The user requirements captured for this proposed solution include making the resource accessible and searchable by dataset, variable and outcome measure. This proposed solution closely aligns with the role of HDR UK's Gateway[8], which enables researchers to discover and request access to UK health-related datasets. We will therefore investigate extending the Gateway's functionality to meet the requirements of the clinical trials community.

Several challenges relate to procedures or processes controlled by data providers, for example obtaining accurate, consistent and timely cost estimates. While many of the proposed solutions are out of scope for the BHF DSC to implement directly, this exercise has generated solutions that could potentially be adopted by data providers. The information captured also provides insights on the needs and priorities of users and may highlight potential solutions that could be both achievable for the data provider and acceptable to meet the needs of users. For example, could research ready datasets to meet common needs be made available at fixed costs? One role of the BHF DSC will be to collate and communicate the needs of users to data providers, and to initiate discussions with the wider group of stakeholders to influence change. As an initial step we will make this report publicly available, including all statements captured in the development of shared solutions (Appendix E, Section 8, Additional material) and will highlight these suggestions to data providers.

In addition to identifying challenges and developing solutions, this work has brought together the community and built consensus around the shared goal of increasing the accessibility and usability of healthcare systems data for cardiovascular clinical trials. It has highlighted the advantages of developing shared solutions, the need to work together as a community to implement these, and the importance of the BHF DSC and HDR UK in coordination.

5 Next steps

The insights from the survey, workshop and wider discussions with the community are already being used to inform the BHF DSC's plans in this area. These plans have been designed to be scalable, generic and of value across the entire cardiovascular trials community. The BHF DSC's aim is to facilitate the use of healthcare systems data in cardiovascular clinical trials, thereby increasing the number and proportion of trials that use such data. We plan to do this through the following initiatives:

- Providing guidance
- Developing best practice
- Standardising clinical outcomes for CVD trials using healthcare systems data
- Using leverage to improve processes
- Targeted driver project activity

To ensure that the views and interests of patients and the public are incorporated into our plan we will seek advice from the BHF DSC's panel of public and patient representatives on where and how this should be incorporated. Areas likely to be the focus of initial input are:

- Obtaining input from patients and public to support our calls for changes to improve the accessibility and usability of healthcare systems data in cardiovascular clinical trials
- Including patient and public representatives in discussions with wider stakeholder groups

- Obtaining input from patients and the public in the prioritisation of outcome measures for the development of standard algorithms/code lists

5.1 Providing Guidance

To help trials teams navigate the process of gaining access to, and using, healthcare systems data by providing independent information and guidance through a suite of webpages on the BHF DSC website, and a "support package" of webinars and open clinics. As an initial step towards making clear what datasets are available and from what providers, we will carry out a scoping exercise to assess the current state of the field. We will also explore extending the functionality of the Gateway to meet the clinical trials community health dataset discoverability and accessibility needs, which would add value to benefit the entire research community.

5.2 Developing best practice

To enable trialists to learn from clinical trials that have successfully used healthcare datasets we will develop a catalogue to bring together and openly share clinical trial case studies. To ensure the catalogue meets the needs of clinical trialists, both in the data it contains and functionality, it will be developed using user-centred design. As an initial step we will identify target users and their needs. We will also develop a community agreed standard template to capture the key characteristics of the clinical trial and healthcare systems dataset access and usage. This standard template will be used to produce the first batch of case studies for storage in the proposed catalogue, and as evidence to encourage trials teams to submit case studies.

We will publish best practice guidelines for access and use of healthcare systems data in clinical trials, drawing on what we have learnt from case studies, the "support package" and driver project activity.

5.3 Standardise clinical outcomes for CVD trials using healthcare systems data

We will work with the community to agree best practices for the derivation, format, and storage of algorithms/code lists to derive clinical trial outcome measures from healthcare systems data. These best practices will be used to create and make available community agreed algorithms/code lists for outcome measures commonly used in cardiovascular clinical trials from healthcare systems data.

5.4 Using leverage to improve processes and data

We will leverage the BHF DSC's role as a community champion to influence changes that improve the accessibility and usability of healthcare systems data for cardiovascular clinical trials. This will include ensuring that we are part of discussions involving the wider clinical trials community, data providers and funders. We will act as a spokesperson for the cardiovascular clinical trials community to ensure their needs are considered.

5.5 Targeted driver project activity

We will support driver projects that are designed to gather evidence regarding the utility of healthcare systems data to successfully support clinical trials

6 Acknowledgements

This workshop and report were produced by the BHF DSC, with the following contributions: Jackie MacArthur co-designed and organised workshop, drafted and revised the report; Matt Sydes co-designed and chaired workshop, and revised the report; Cathie Sudlow provided input on workshop design, co-chaired, and revised the report; Gavin Murphy provided input on workshop design; Lynn Morrice provided input on workshop design, and revised the report; Tammy Watchorn (https://tammywatchorn.com/) designed workshop breakout sessions; Lydia Martin and Sue Page provided administrative and organisational support.

In addition, the National Cardiac Surgery Clinical Trials Initiative members, workshop participants (see Appendix A) and survey respondents are acknowledged for their engagement and rich contributions.

This work was funded by the BHF as part of the BHF Data Science Centre led by HDR UK (BHF Grant no. SP/19/3/34678). The National Cardiac Surgery Clinical Trials Initiative is funded by Heart Research UK and The Royal College of Surgeons.

7 References

1. Martin, L., Hutchens, M., Hawkins, C. & Radnov, A. How much do clinical trials cost? *Nat. Rev. Drug Discov.* **16**, 381–382 (2017).
2. Moore, T. J., Heyward, J., Anderson, G. & Alexander, G. C. Variation in the estimated costs of pivotal clinical benefit trials supporting the US approval of new therapeutic agents, 2015-2017: a cross-sectional study. *BMJ Open* **10**, e038863 (2020).
3. Lensen, S. *et al.* Access to routinely collected health data for clinical trials - review of successful data requests to UK registries. *Trials* **21**, 398 (2020).
4. MacArthur, J., Morrice, L., Sudlow, C. & Sydes, M. BHF Data Science Centre: Data-Enabled Trials Survey Report (0.1). *Zenodo* doi:https://doi.org/10.5281/zenodo.5079178.
5. National Cardiac Surgery Clinical Trials Initiative. https://le.ac.uk/cardiovascular-sciences/about/heart-surgery/national-cardiac-surgery-clinical-trials-initiative.
6. Challenge statements gathered during BHF DSC cardiac surgery workshop. https://www.hdruk.ac.uk/wp-content/uploads/2022/03/BHF_DSC_Cardiac_surgery_workshop_Challenge_statements_gathered.xlsx.
7. Sydes, M. R. *et al.* Realising the full potential of data-enabled trials in the UK: a call for action. *BMJ Open* **11**, e043906 (2021).
8. The Gateway. https://www.healthdatagateway.org/.

8 Additional material

Appendix A – Workshop attendee list[1]

Name	Affiliation
Alexander Perkins	London School of Hygiene and Tropical Medicine
Ana Suazo Di Paola	Leicester Clinical Trials Unit, University of Leicester
Ayesha Mathias	Newcastle Clinical Trials Unit
Barnaby Reeves	University of Bristol
Ben Gibbison	University of Bristol
Carla Richardson	Leicester Clinical Trials Unit, University of Leicester
Cassandra Brookes	Leicester Clinical Trials Unit, University of Leicester
Catherine Wloch	UK Health Security Agency
Cathie Sudlow	BHF Data Science Centre
Chris Gale	University of Leeds
Daniel Fudulu	University of Bristol
Debbie Ringham	BHF Data Science Centre
Edmund Wyatt	Surgical Interventions Trials Unit, University of Oxford
Enoch Akowuah	South Tees NHS Foundation Trust, Newcastle University
Fang Smith	University of Birmingham
Florence Lai	University of Leicester
Gavin Murphy	University of Leicester
Gerry McCann	University of Leicester
Gianni Angelini	University of Bristol
Gudrun Kunst	King's College London
Hardeep Aujla	University of Leicester
Iain Squire	University of Leicester
Jackie MacArthur[B]	BHF Data Science Centre
Jeremy Dearling[P]	Public representative
Joanne Miksza	University of Leicester
John Cleland	University of Glasgow
Jonathan Prichard	Newcastle clinical trials unit
Judith Tanner	University of Nottingham
Leonardo Teixeira[B]	
Linda Sharples	London School of Hygiene and Tropical Medicine
Luke J. Rogers	Bristol Heart Institute
Lydia Martin	BHF Data Science Centre
Lynn Morrice[B]	BHF Data Science Centre
Mahmoud Loubani	Hull University Teaching Hospitals

[1] Includes all participants that have consented to have their name and affiliation made publicly available, in association with this workshop
[B]Breakout session chair/facilitator
[P]Public contributor

Name	Affiliation
Maria Pufulete	University of Bristol
Mario Cibelli	Imperial College London
Marius Roman	University of Leicester
Mark Lewis[P]	Public representative
Mark Petrie	University of Glasgow
Matt Sydes	BHF Data Science Centre, University College London
Melissa Rochon	RBHH/Guy's & St Thomas' NHS FT
Michelle Bardgett	Newcastle CTU
Mirjana Sirovica	University of Leicester Clinical Trials Unit
Nadjat Medeghri	Surgical Intervention Trials Unit, University of Oxford
Nafisa Boota	University of Leicester
Paul Maslowski	Leicester Clinical Trials Unit
Phil Blakelock[P]	Public representative
Philippa Watts	Newcastle University
Rachel Perry	University of Bristol
Ralph Stewart	Auckland City Hospital, NZ
Rana Sayeed	Oxford University Hospitals NHS Foundation Trust
Rebecca Maier	Newcastle Clinical Trials Unit
Ricky Vaja	Imperial college London
Roopa Rajagopal	University of Leicester
Rosalie Magboo	Barts Health NHS Trust and Queen Mary University of London
Ruth Knight	OCTRU, CSM, University of Oxford
Sarah Murray[P]	Public representative
Saul Myerson	Radcliffe Department of Medicine, University of Oxford
Sian Baldock	Leicester Clinical Trials Unit
Sue Page	University of Leicester
Suraj Pathak	University of Leicester
Tammy Watchorn[B]	
Theresa Lamagni	UK Health Security Agency
Thomas Treibel	Barts Heart Centre and University College London
Tim Clayton	London School of Hygiene & Tropical Medicine Clinical Trials Unit
Tim Dong	University of Bristol
Victoria Wheeldon	Newcastle Clinical Trials Unit
Yvette Walters	University of Leicester

Appendix B – Agenda

Day 1 – Friday 8th October 2021

All times are in BST

Time	Session 1: Setting the scene (chair Cathie Sudlow)	Speakers
9:00	Welcome and introduction to the workshop	Matt Sydes
9:05	Introduction to the BHF Data Science Centre	Cathie Sudlow
9:15	The use of routinely-collected health data in clinical trials	Matt Sydes
9:25	Challenges survey results	Jackie MacArthur
9:30	Introduction to the National Cardiac Surgery Clinical Trials Initiative	Gavin Murphy
9:35	Lightning talks from cardiac surgery trials (see below)	Cardiac surgery trials' leads
10:10	Questions	
10:25	Introduction to breakout sessions	Tammy Watchorn
10:30	Break	
Time	**Session 2: Interactive breakout sessions to gather input on the challenges faced in accessing and using routinely-collected data in cardiac surgery trials (sessions coordinated by Tammy Watchorn)**	
	Groups A1 and A2	**Groups B1 and B2**
10:45-11:45	Breakout discussion	*(Free time)*
12:00-13:00	End of workshop 1	Breakout discussion
13:00		End of workshop 1

Day 2 – Friday 15th October 2021

All times are in BST

Time	Session 3: Interactive breakout session developing shared solutions to the challenges identified in workshop 1 (sessions coordinated by Tammy Watchorn)	
	Groups A1 and A2	**Groups B1 and B2**
9:00-10:15	Breakout discussion	*(Free time)*
10:30-11:45	*(Free time)*	Breakout discussion
11:45	Break	
Time	**Session 4: Reporting back and next steps (Chair Matt Sydes)**	**Speakers**
12:00	Reporting back	Chair/facilitator from each breakout group reports back
12:30	Chaired discussion and next steps	Cathie Sudlow and Matt Sydes
13:00	End of workshop 2	

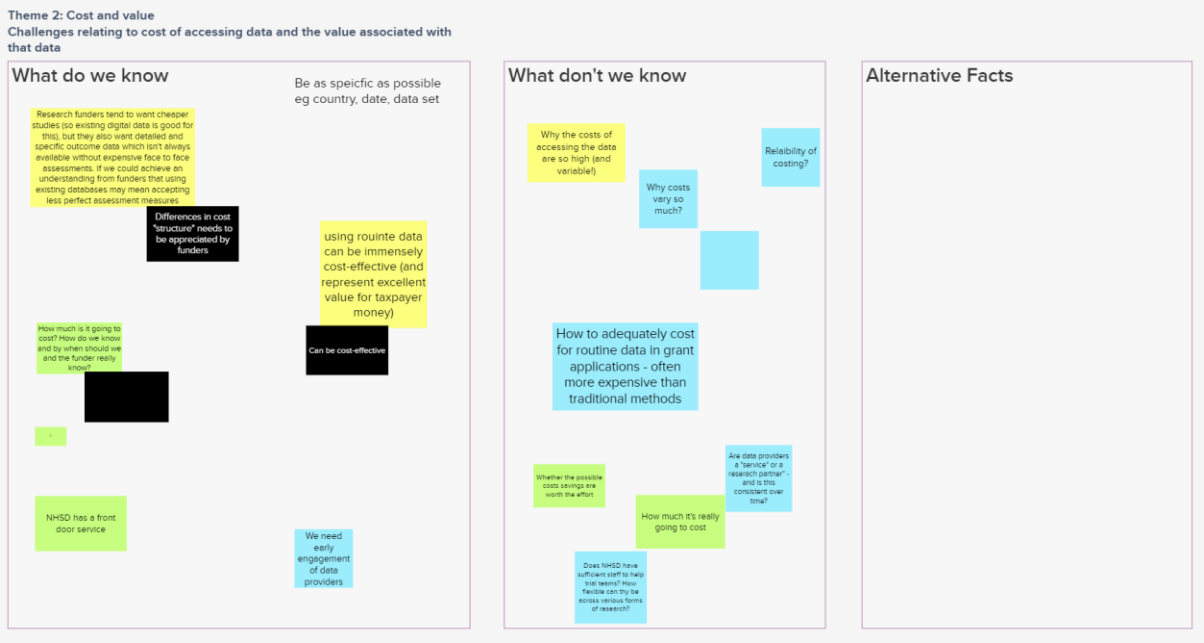
Lightning talks from cardiac surgery trials

Title	Presenter	Clinical Study Group
Chronic post-sternotomy pain trial and UKRI Datahub	Fang Gao Smith	1 (Long Term Outcomes and Quality of Life)
REVASC-HF-UK	Mark Petrie	1 (Long Term Outcomes and Quality of Life)
Using routine data to characterise the cardiac surgery population and identify inequality	Maria Pufulete	2 (Prehabilitation)
UK Asymptomatic Mitral Regurgitation Study (UK-AdMiRe)	Rana Sayeed and Saul Myerson	3 (Heart Valve Interventions)
EASY-AS	Gerry McCann	3 (Heart Valve Interventions)
PROTECT MVR	Enoch Akowuah	4 (Minimally Invasive, Hybrid and Percutaneous Techniques)
Volat**I**le **V**ersus **I**ntravenous **AN**esthesia in Cardiac Surgery (**VIVIAN**) - a Multicentre Randomised Controlled Trial	Gudrun Kunst	5 (Organ Protection)
Role of hypothermia in adult cardiac surgery patients	Gudrun Kunst	5 (Organ Protection)
Target Wound Infection	Luke Rogers and Ricky Vaja	7 (Infection Prevention)
The impact of frailty and inflammaging in patients with Acute Coronary Syndrome	Marius Roman	9 (Frailty, Sarcopenia and Chronic Conditions)

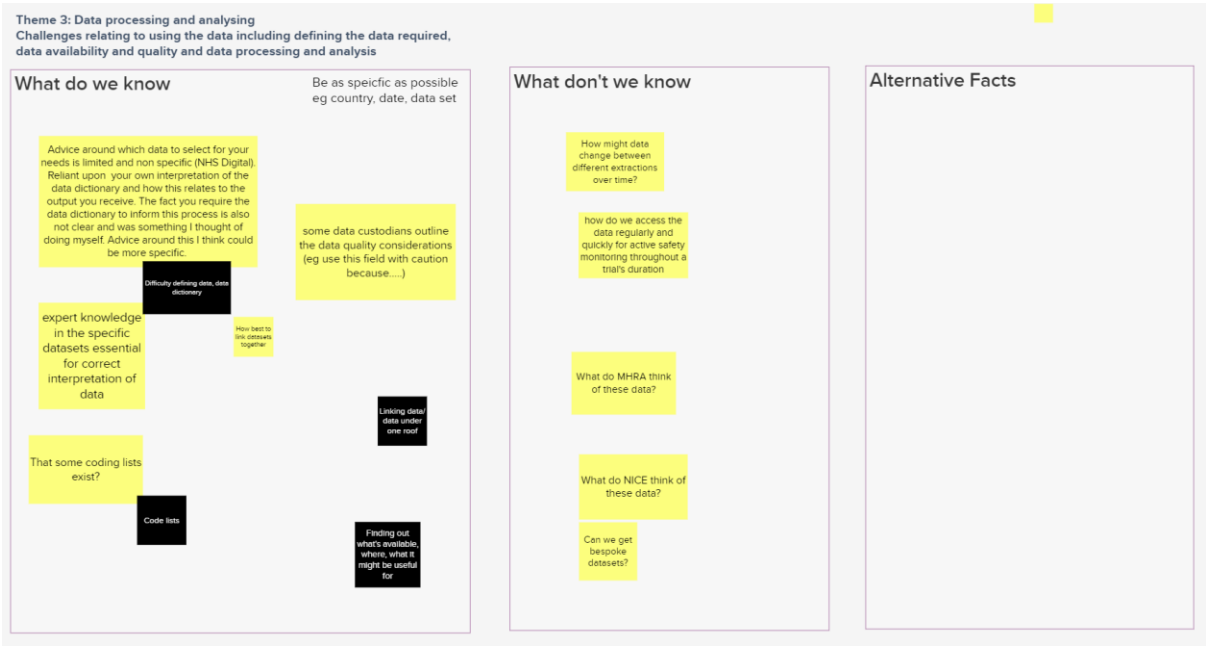
Appendix C - Mural board images from workshop part 1 "Defining challenges"

Group A1 – Challenges associated with gaining approval/access

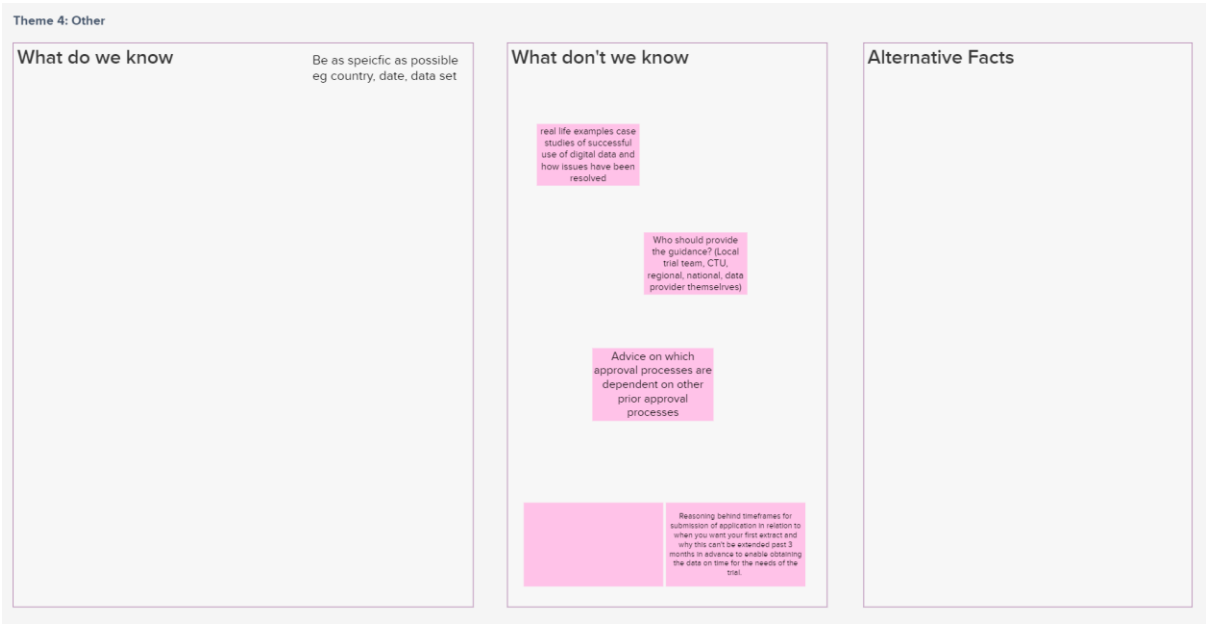
Group A1 – Challenges related to the cost of accessing data and the value of that data



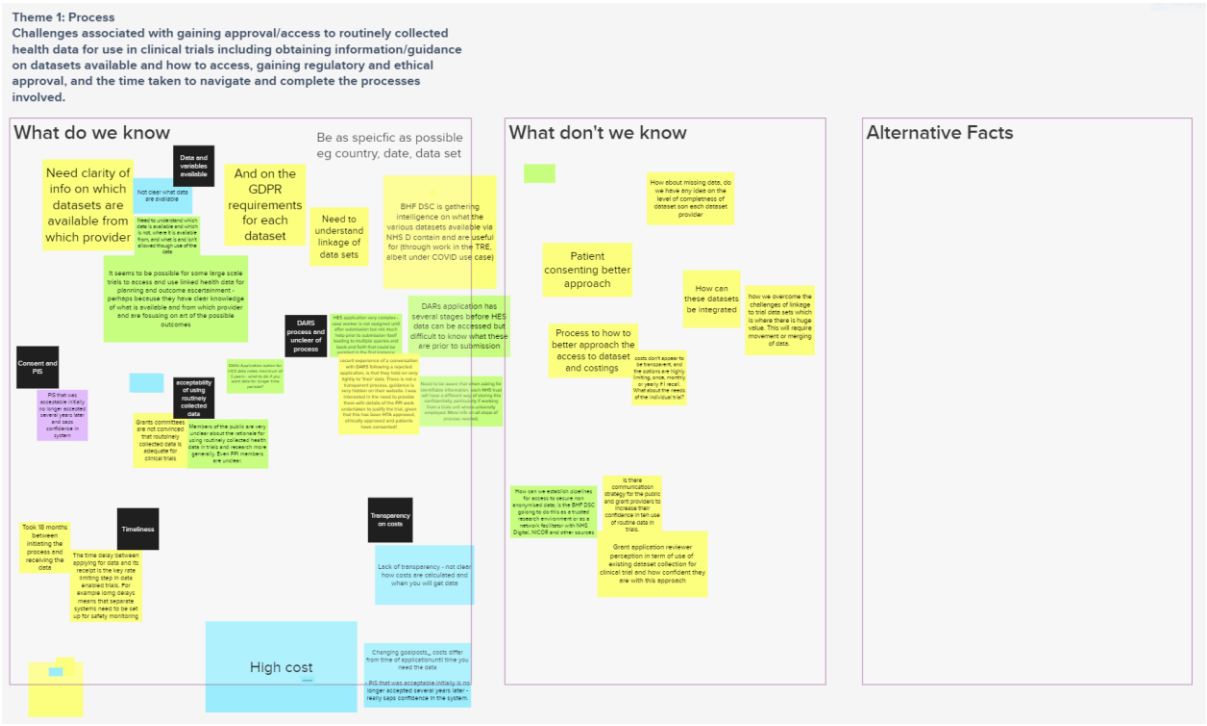
Group A1 – Challenges associated with data processing and analysing



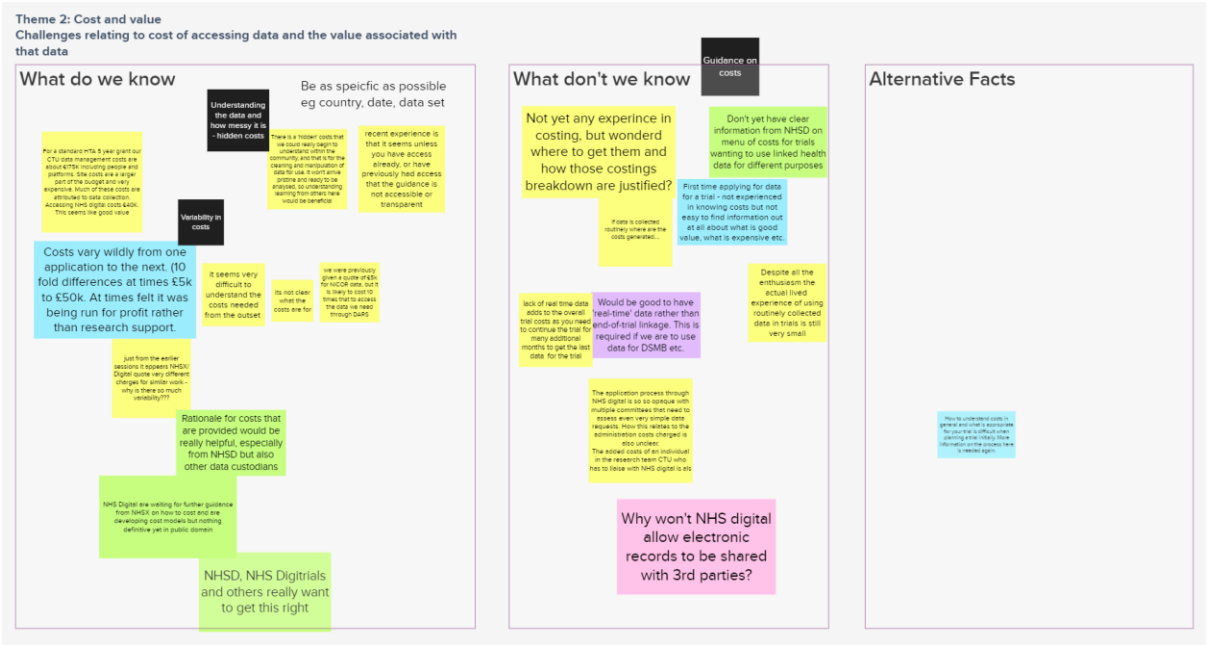
Group A1 – Other challenges



Group A2 – Challenges associated with gaining approval/access

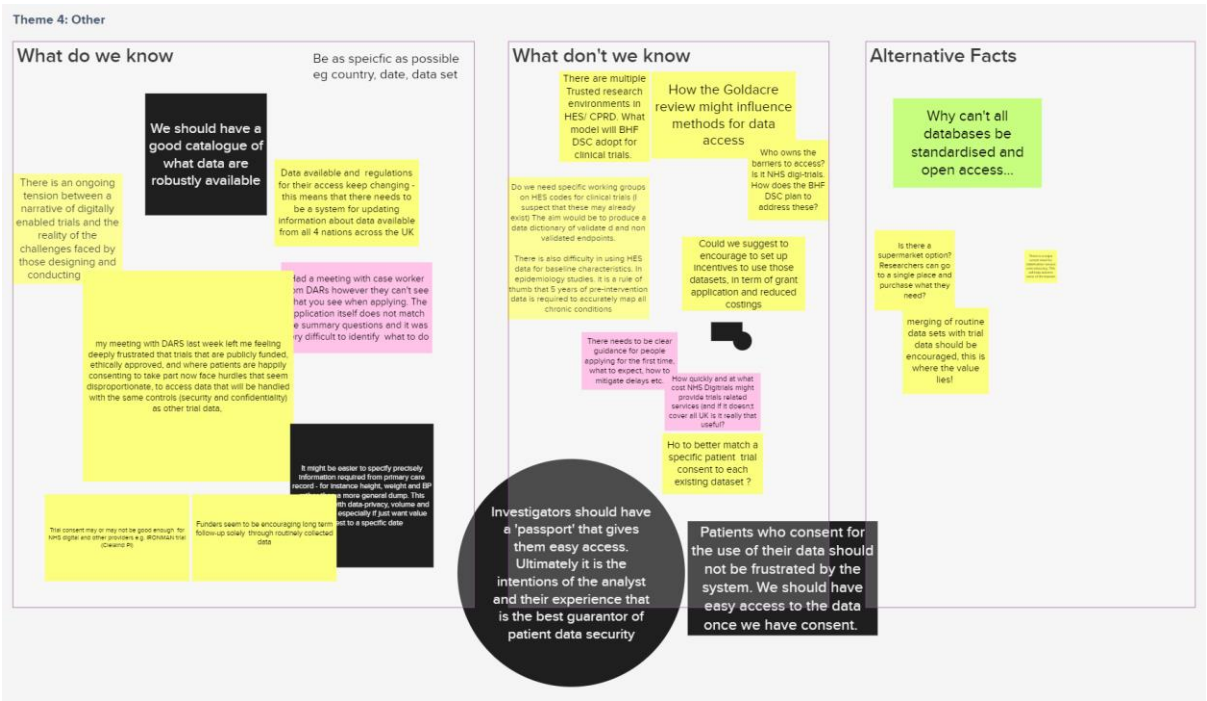


Group A2 – Challenges related to the cost of accessing data and the value of that data

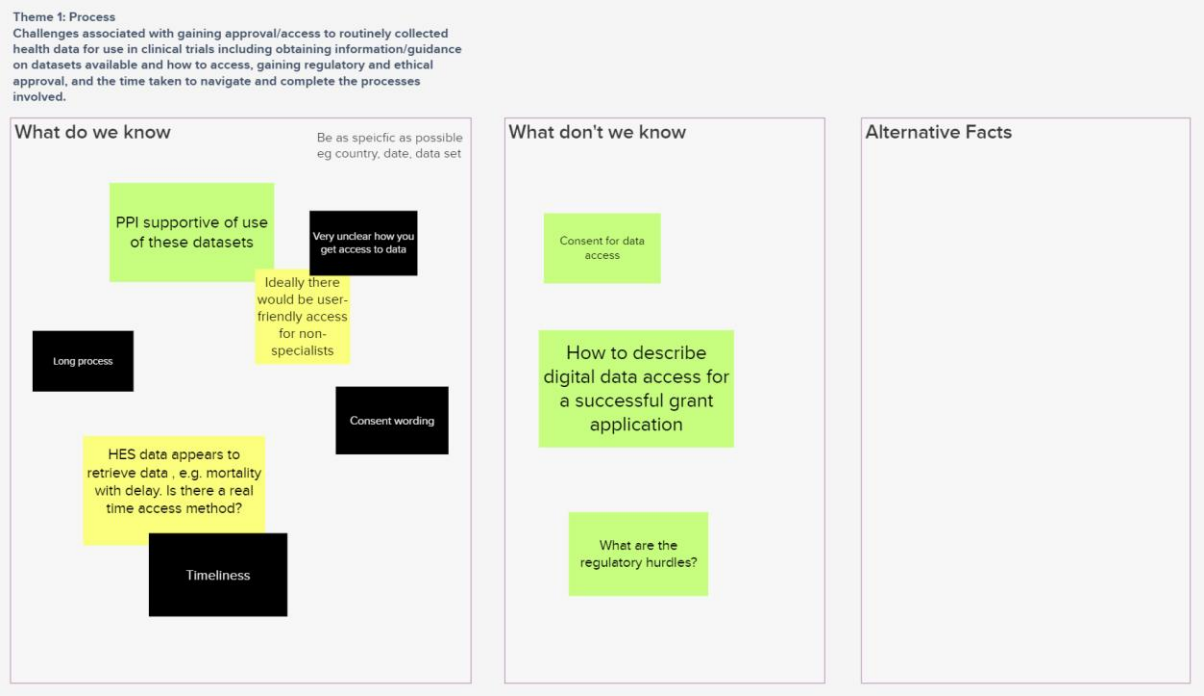


Group A2 – Challenges associated with data processing and analysing

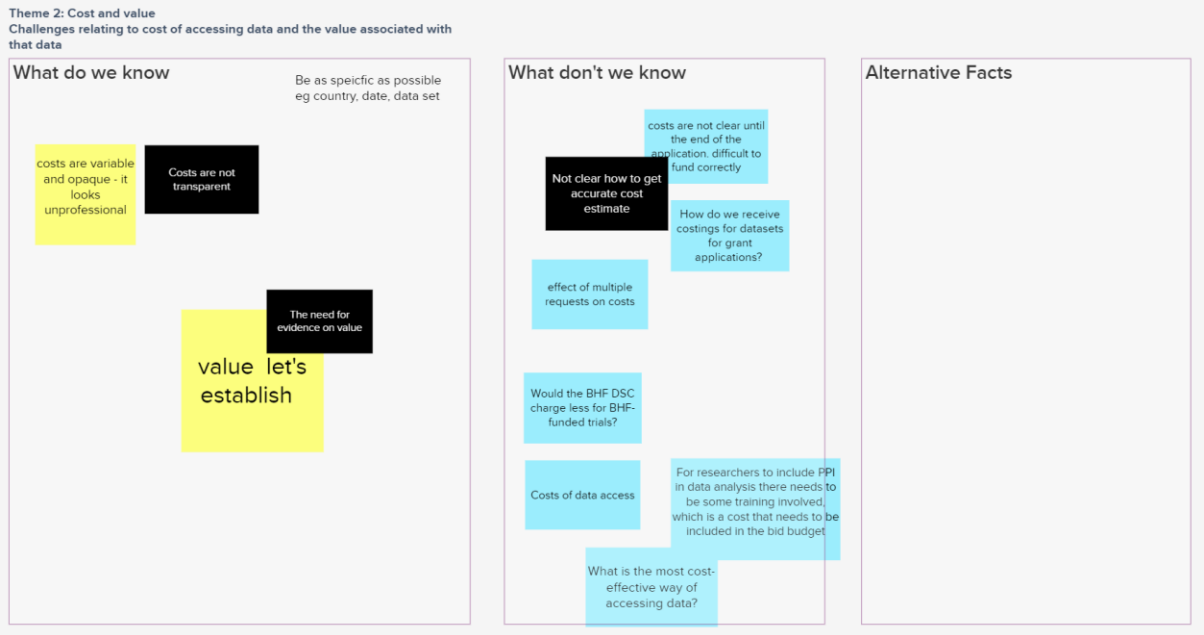
Group A2 – Other challenges



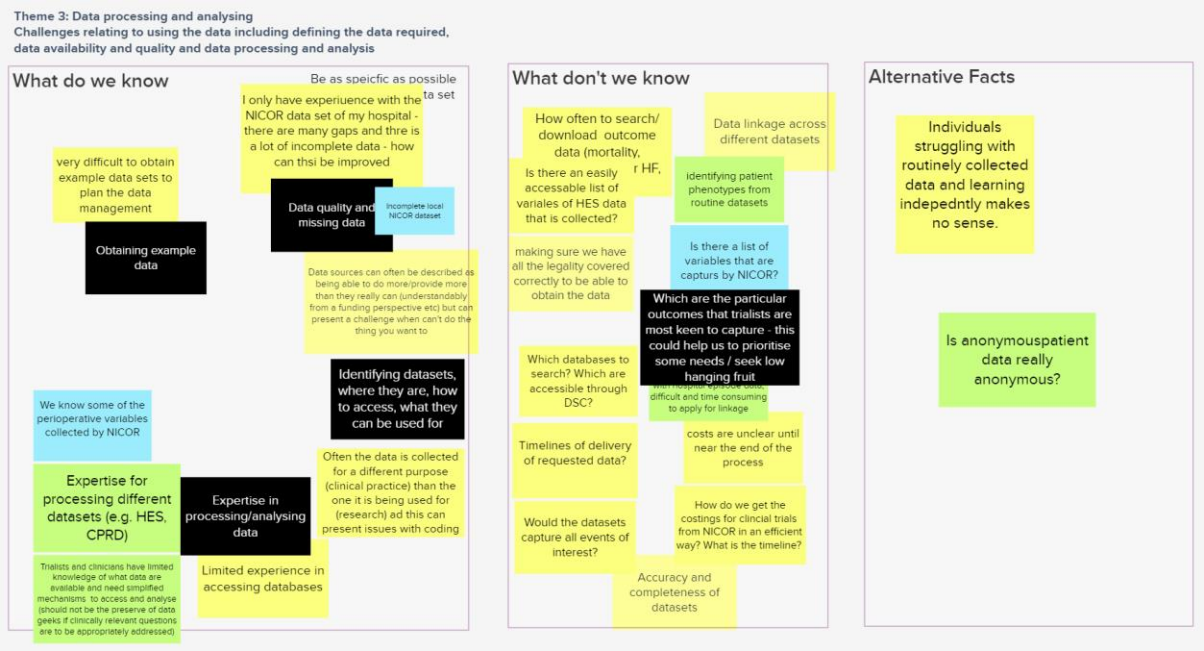
Group B1 – Challenges associated with gaining approval/access



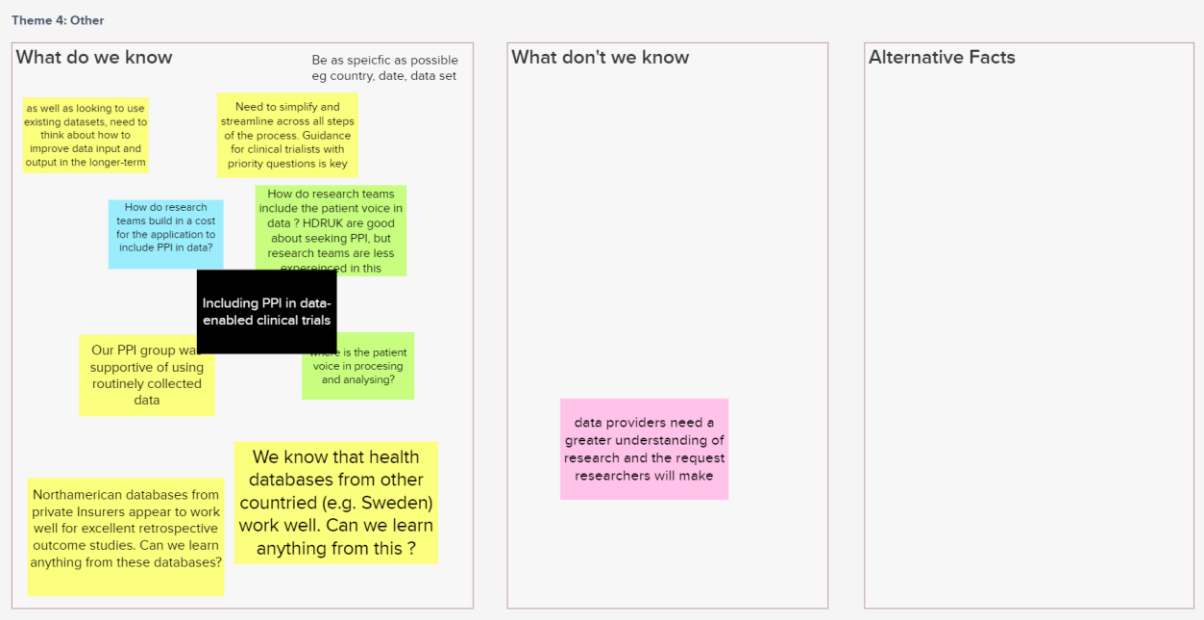
Group B1 – Challenges related to the cost of accessing data and the value of that data



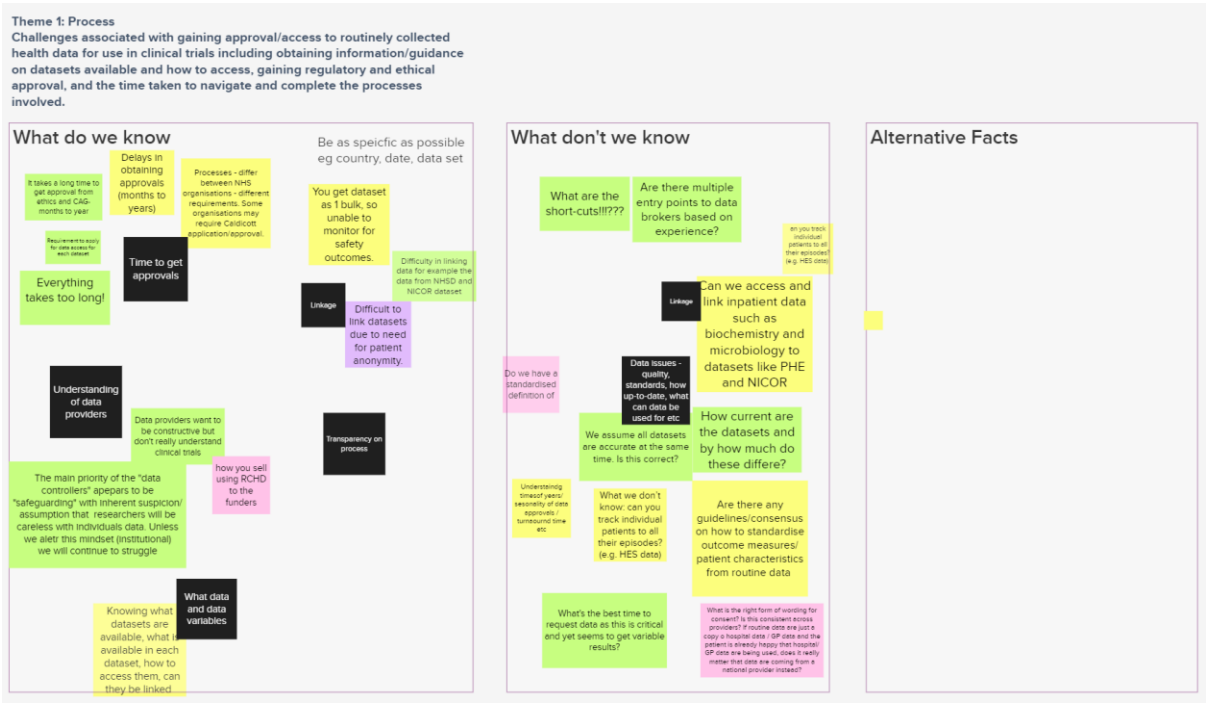
Group B1 – Challenges associated with data processing and analysing



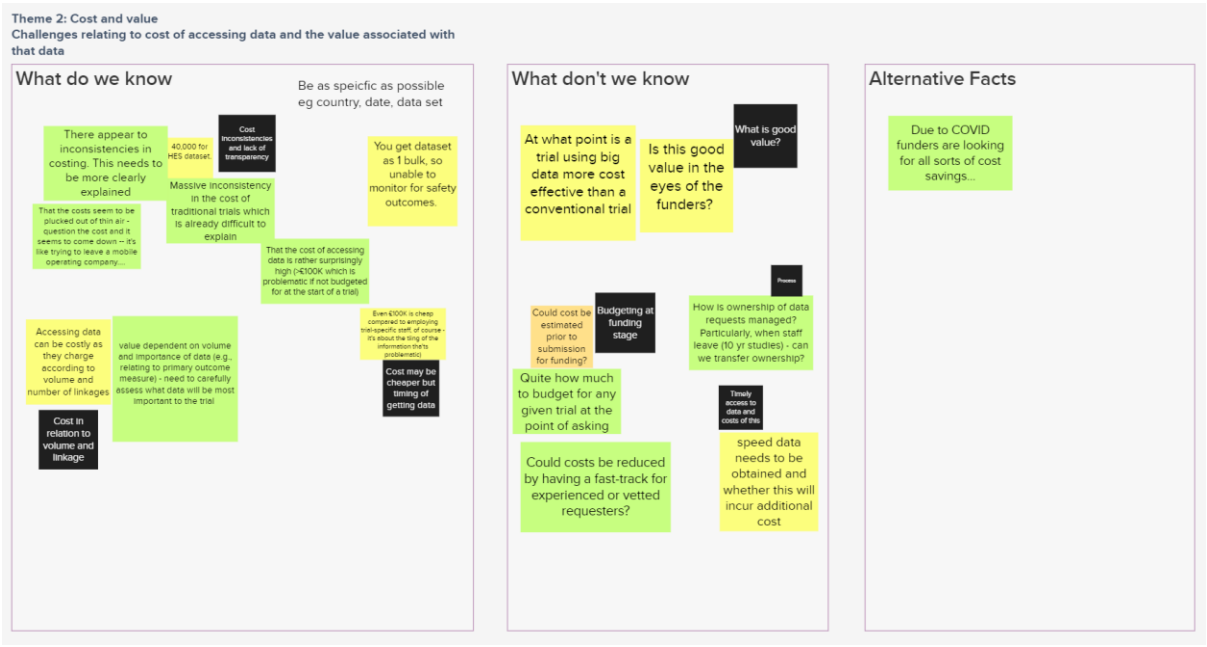
Group B1 – Other challenges



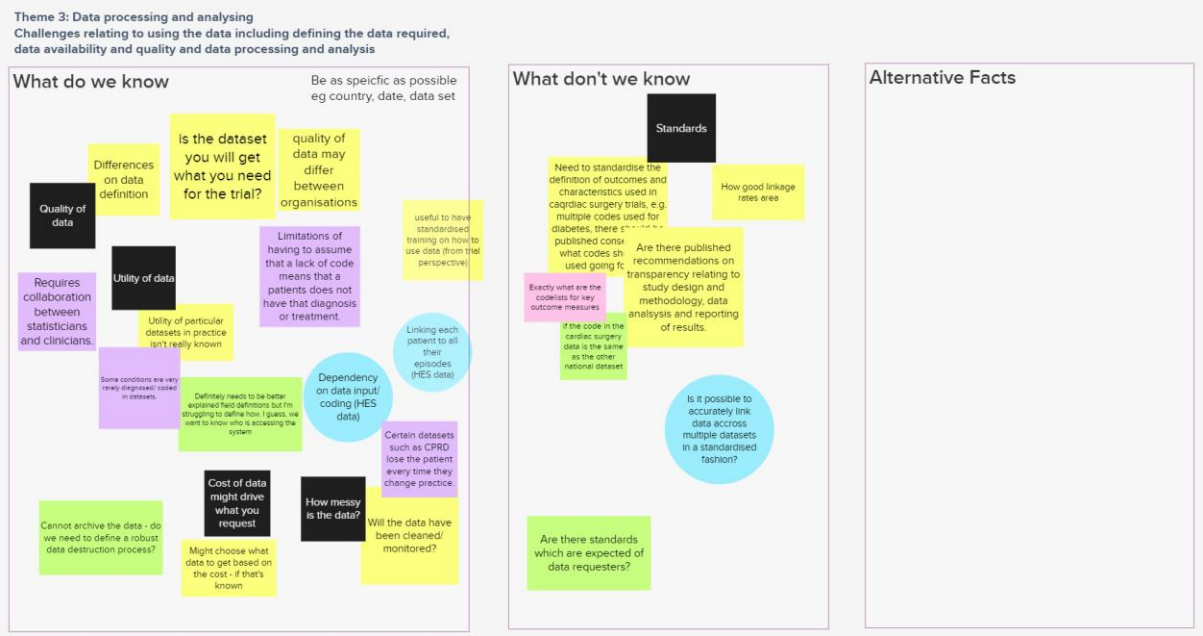
Group B2 – Challenges associated with gaining approval/access



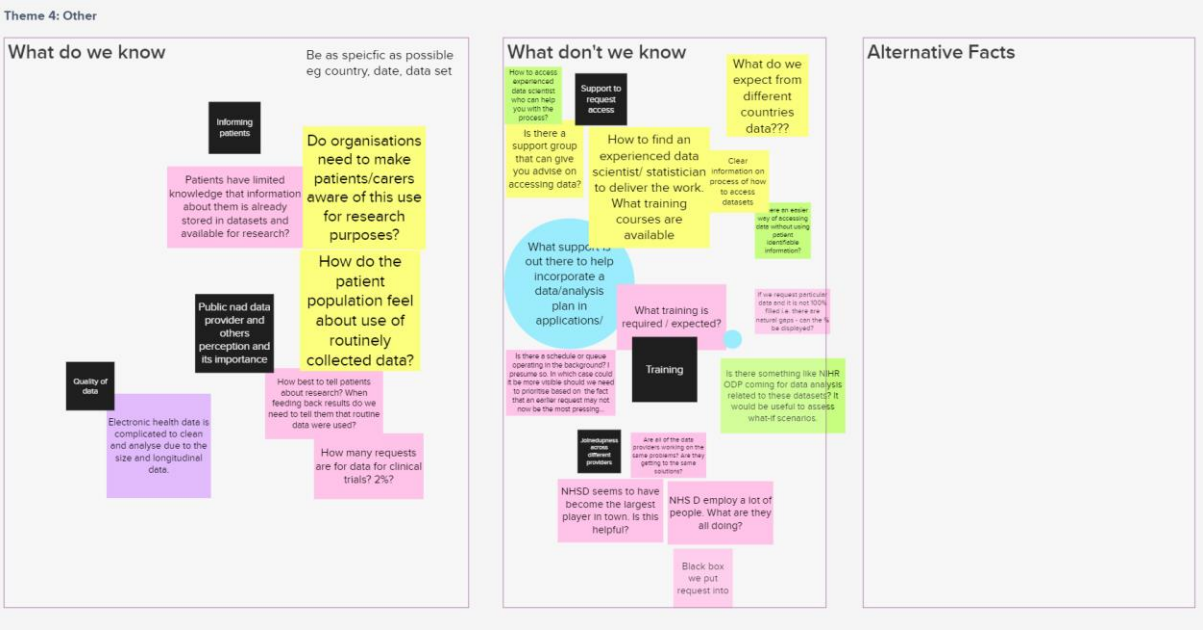
Group B2 – Challenges related to the cost of accessing data and the value of that data



Group B2 – Challenges associated with data processing and analysing

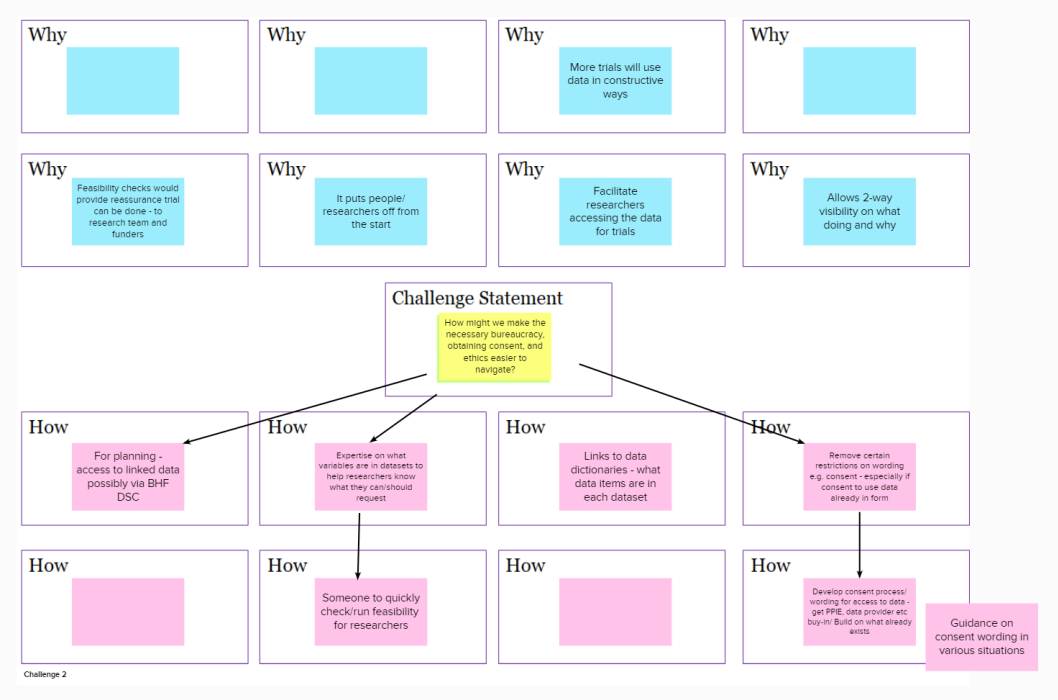


Group B2 – Other challenges

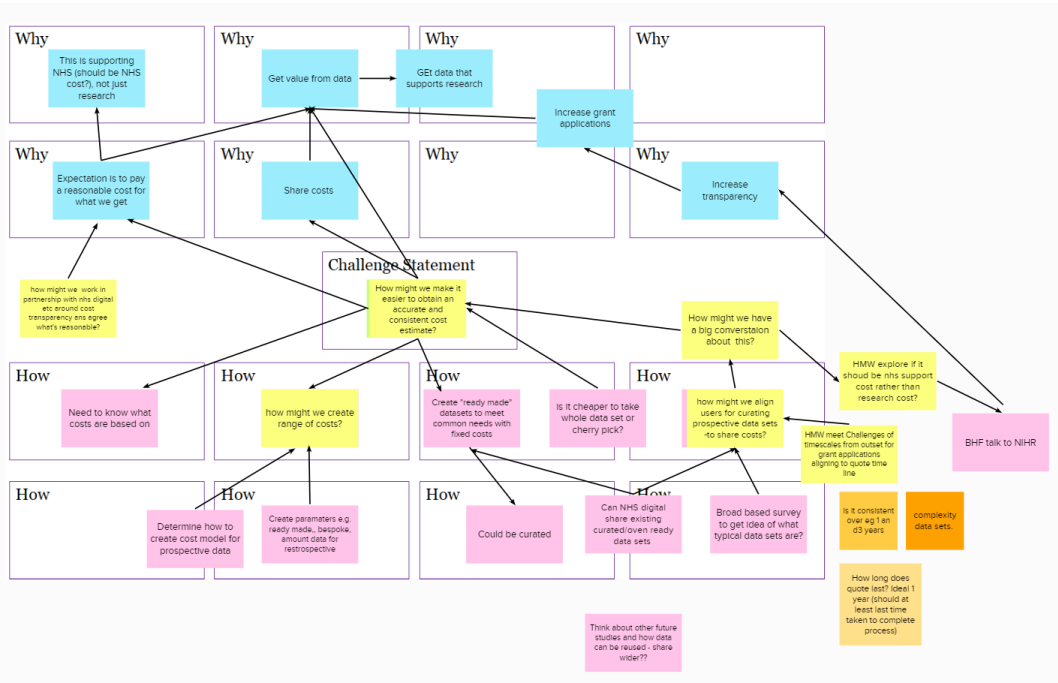


Appendix D - Mural board images from workshop part 2 "Working towards shared solutions"

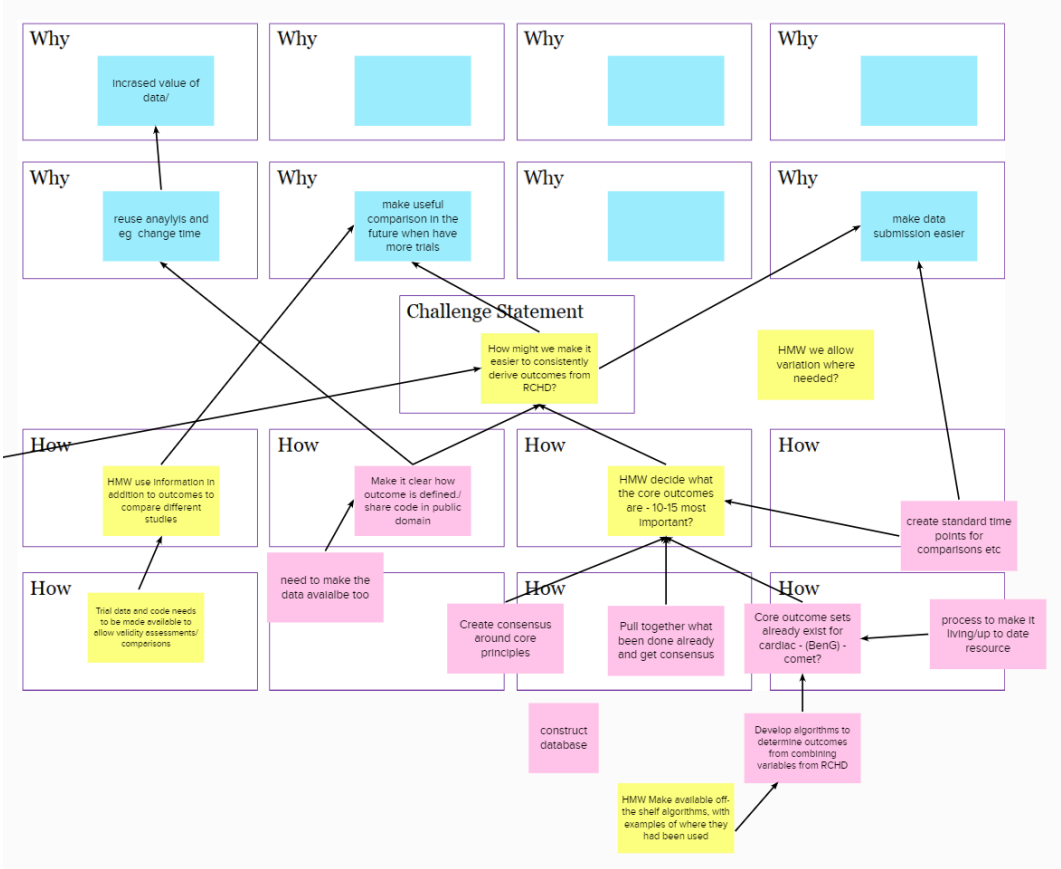
D1 – How might we make the necessary bureaucracy, obtaining consent, and ethics easier to navigate? – Group A1



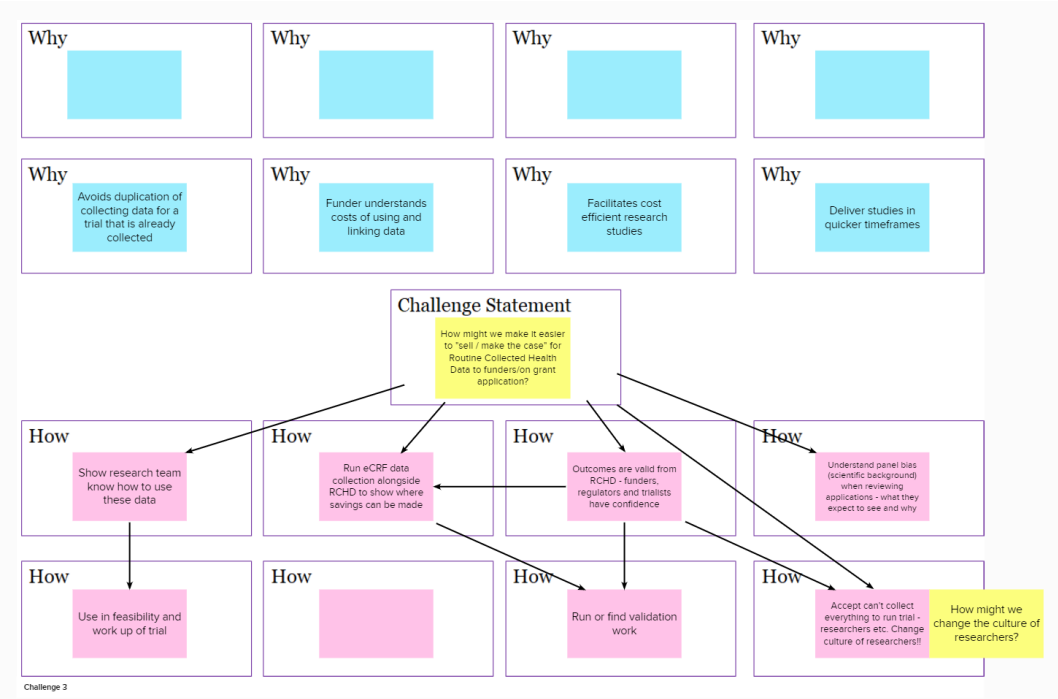
D2 - How might we make it easier to obtain an accurate and consistent cost estimate? – Group A2



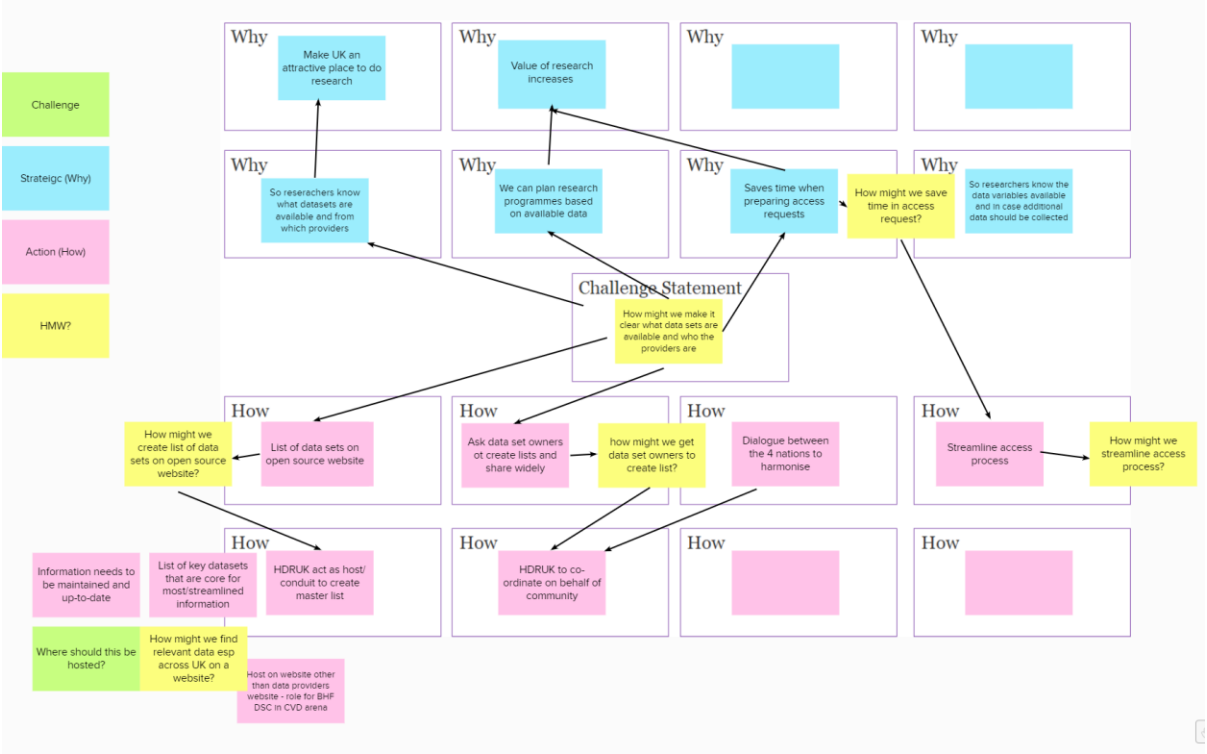
D3 - How might we make it easier to consistently derive outcomes from RCHD? – Group B2



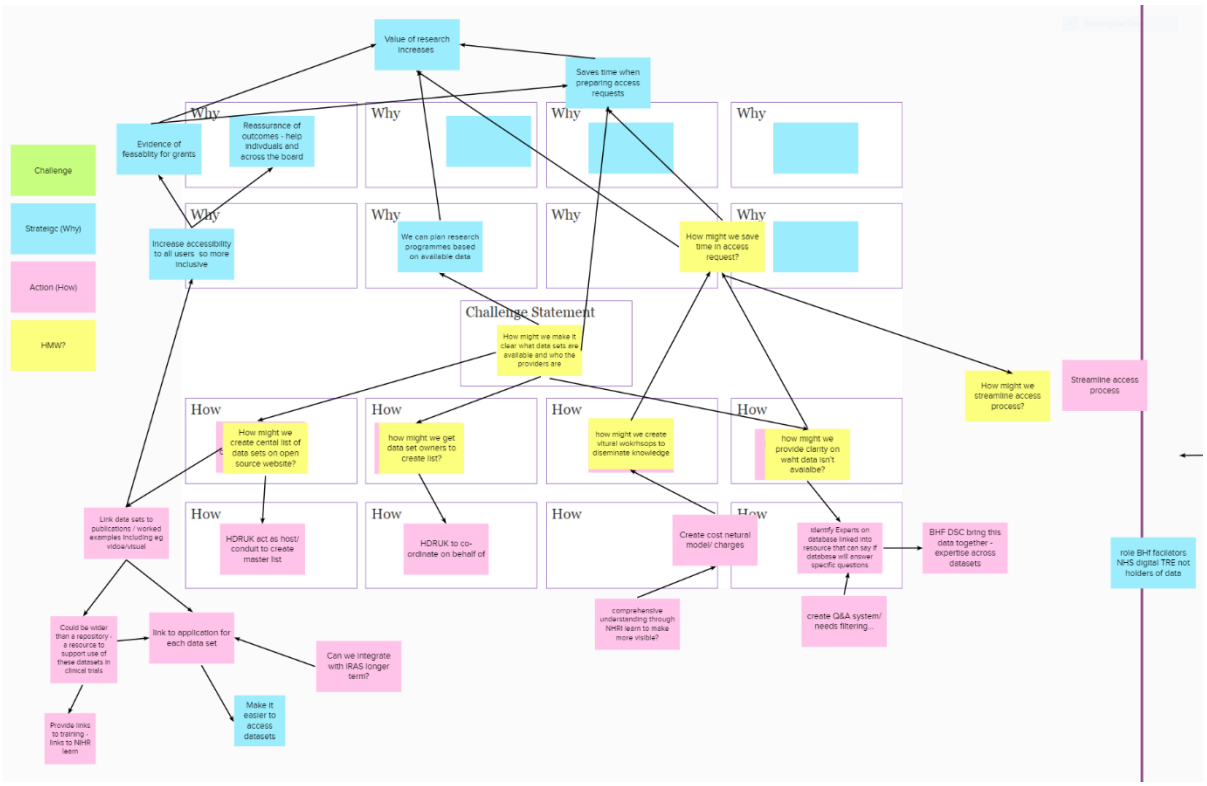
D4 - How might we make it easier to "sell" RCHD to funders/on grant application? – Group A1



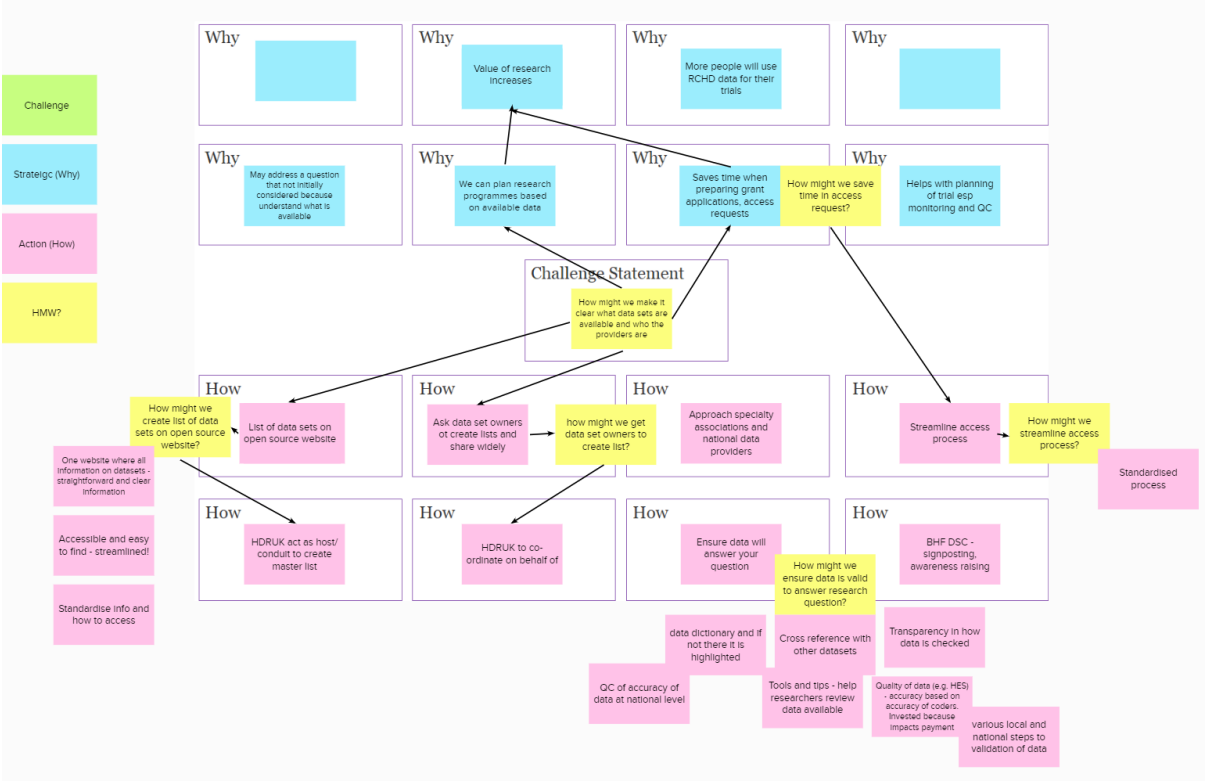
D5 - How might we make clear what datasets are available and from what providers? – Group A1



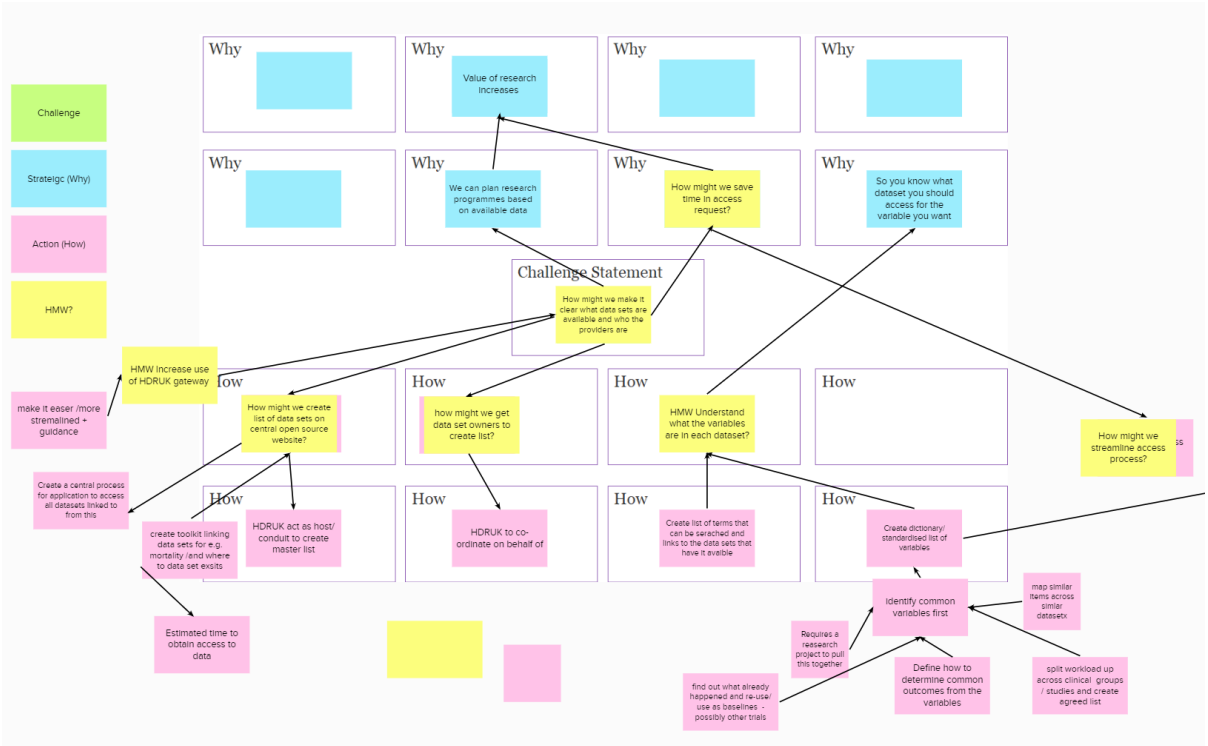
D6 - How might we make clear what datasets are available and from what providers? – Group A2



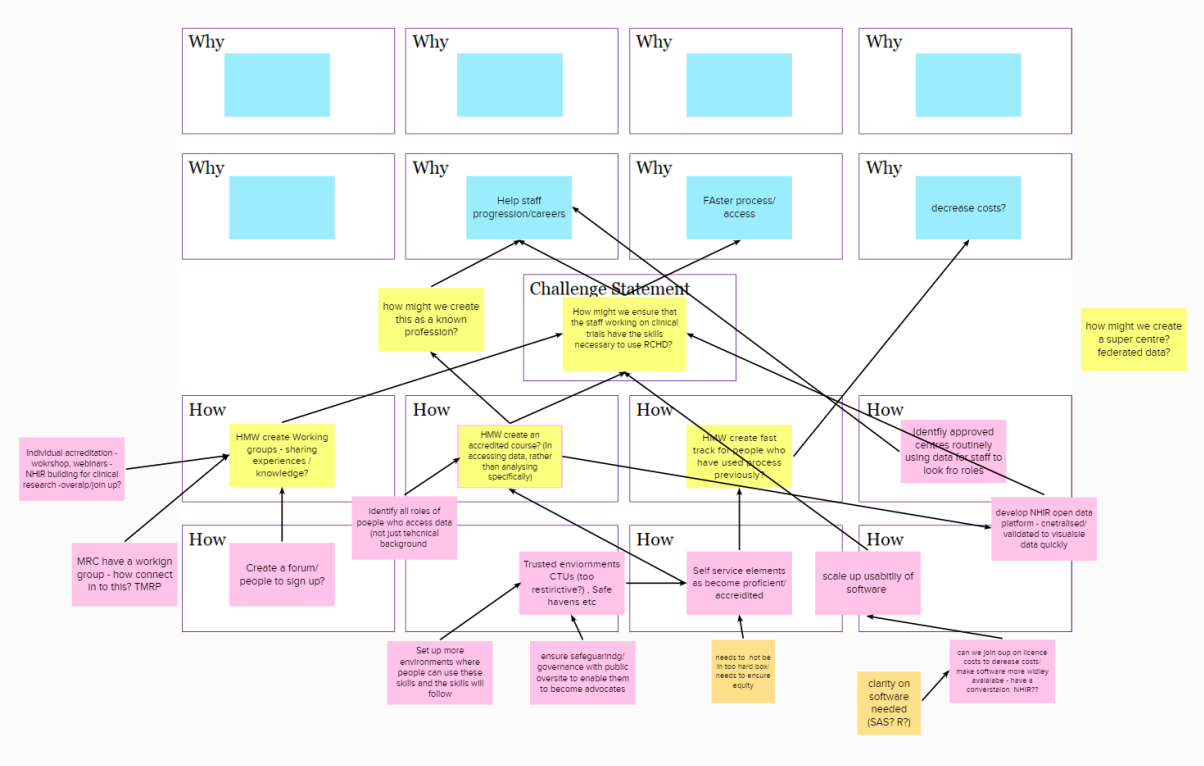
D7 - How might we make clear what datasets are available and from what providers? – Group B1



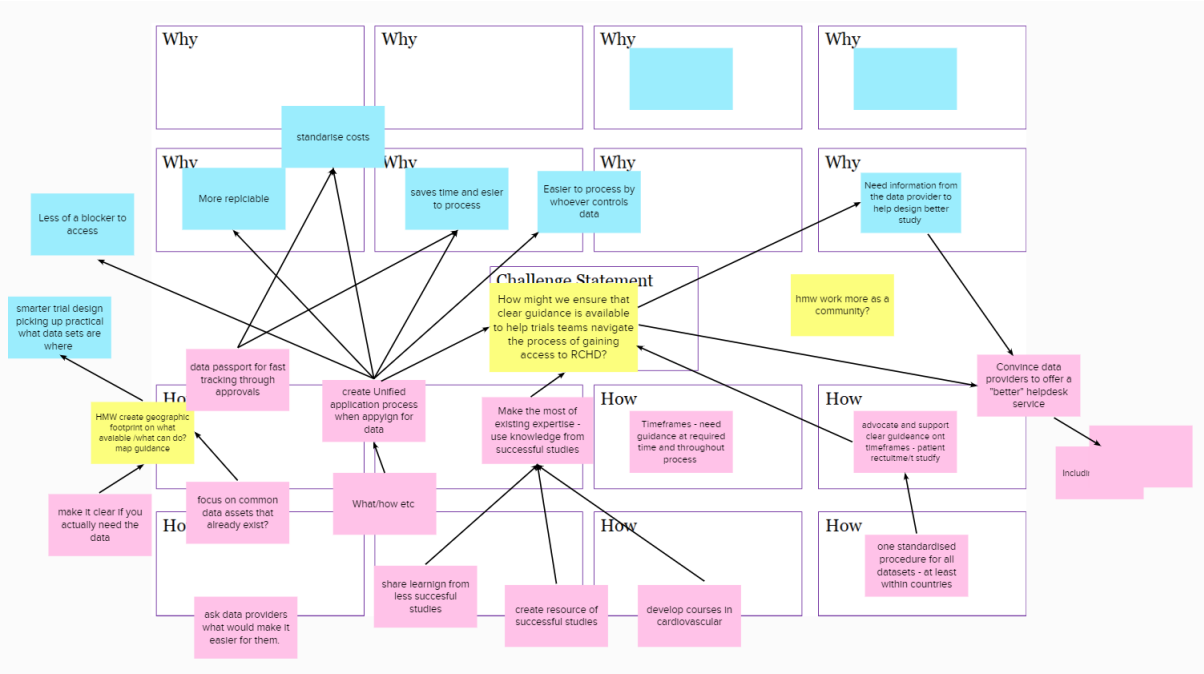
D8 - How might we make clear what datasets are available and from what providers? – Group B2



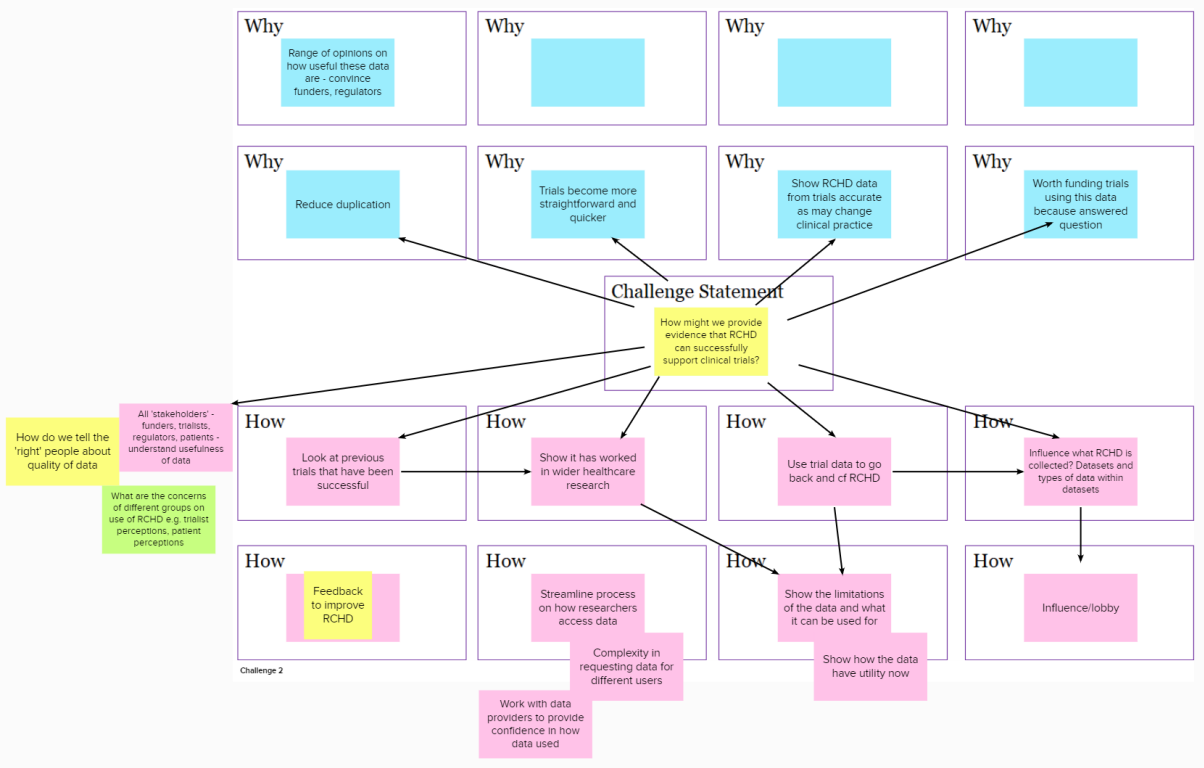
D9 - How might we ensure that the staff working on clinical trials have the skills necessary to use RCHD? – Group A2



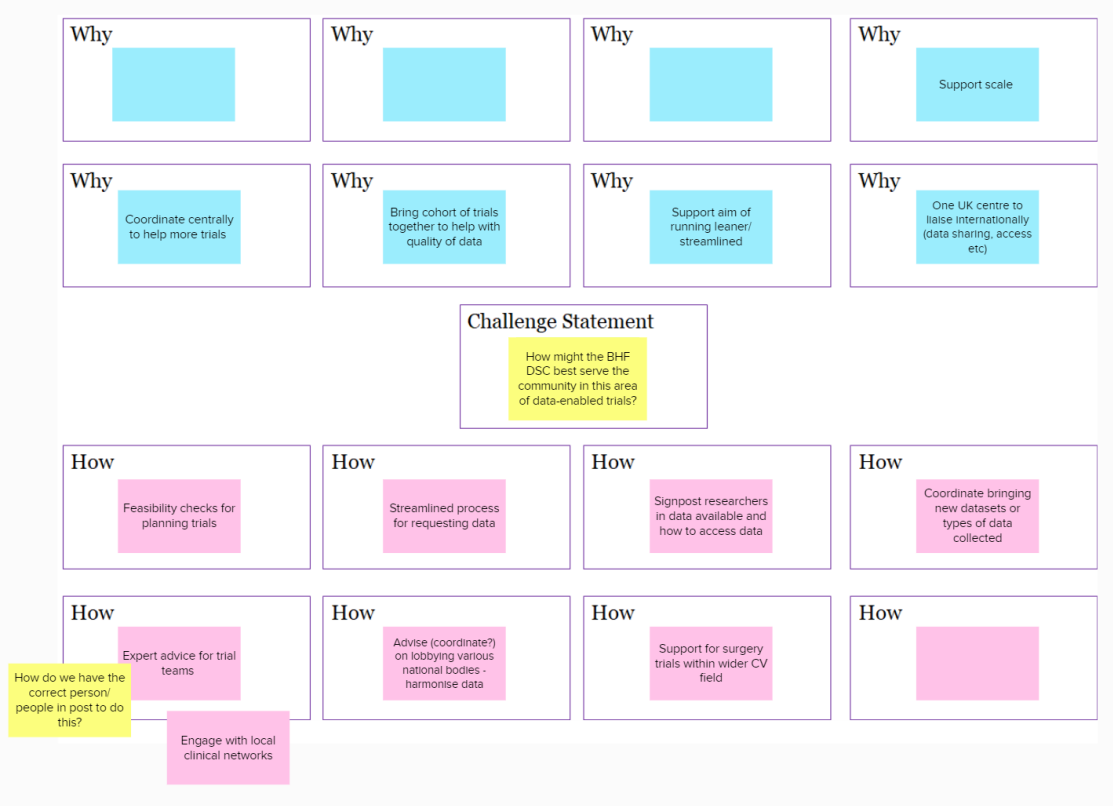
D10. How might we ensure that clear guidance is available to help trials teams navigate the process of gaining access to RCHD? – Group B2



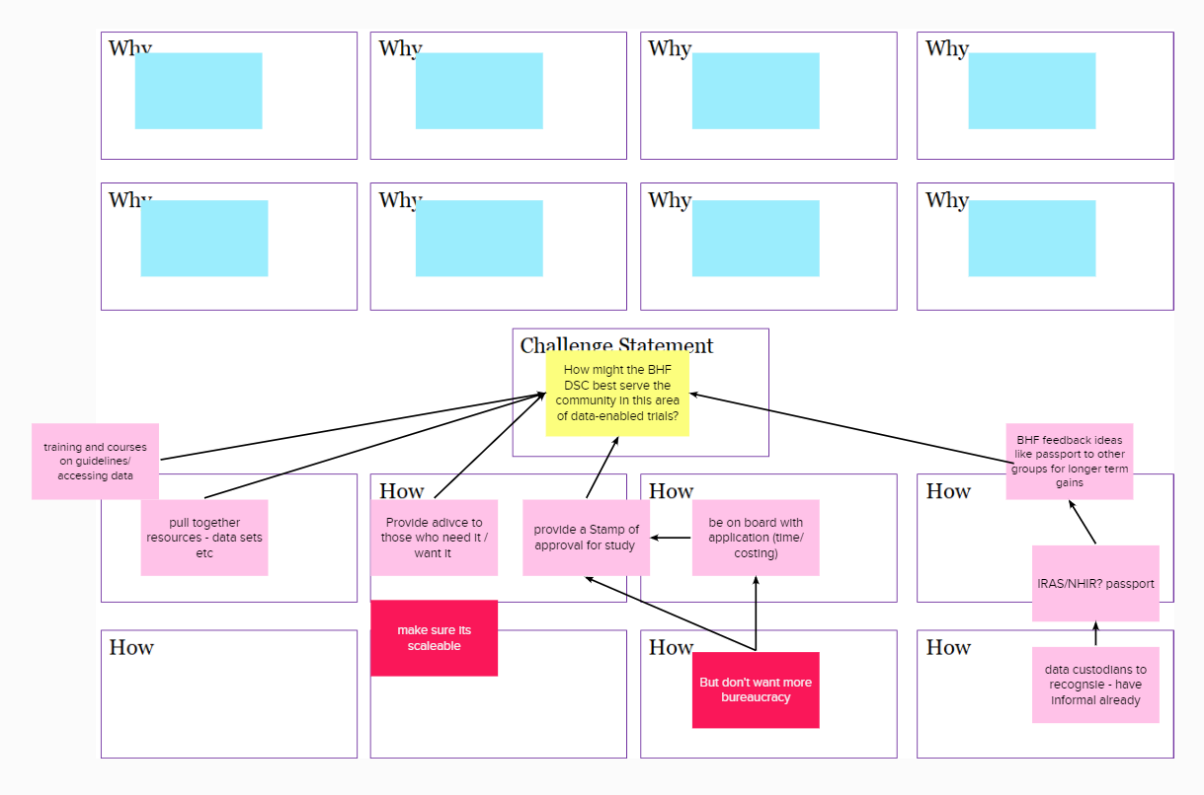
D11. How might we provide evidence that RCHD can successfully support clinical trials? – Group B1



D12. How might the BHF DSC best serve the community in this area of data-enabled trials? – Group B1



D13. How might the BHF DSC best serve the community in this area of data-enabled trials? – Group B2



Appendix E – All statements captured on Mural boards during workshop part 2 "Working towards shared solutions"

How might we make it clear what datasets are available and who the providers are?

Why

- Researchers know and can easily identify what datasets are available and from which providers
    - The UK is an attractive place to do research
- Research can be planned based on available data
- Time is saved when planning research and preparing access requests
    - Value of research increases
- Increase accessibility to all users, so more inclusive
    - Evidence of feasibility for grants
    - Reassurance of outcomes – help individuals and across the board
- May address a question not initially considered because understand what is available
- More people will use RCHD for their trials
- Helps with planning of trial, especially monitoring and QC
- So you know what dataset you should access for the variable you want

How

- Create a list of datasets on open-source website
    - HDR UK acts as host/conduit to create master list
        - Create list of key datasets that are core for most/streamlined information
        - Information needs to be maintained and up-to-date
    - Link datasets to publications/worked examples including e.g. video/visual
        - Could be wider than a repository – a resource to support use of these datasets in clinical trials
            - Provide links to training – links to NIHR learn
            - Link to application for each dataset – make it easier to access datasets
                - Can we integrate with IRAS longer term?
    - One website where all information on datasets – straightforward and clear information
        - Accessible and easy to find – streamlined
        - Standardise info and how to access
    - Create a central process for application to access – all datasets linked from this
    - Create a toolkit linking datasets for e.g. mortality and where where data set exists
        - Estimated time to obtain access to data
    - Include information on what variables are in each dataset
        - Create list of terms that can be searched and link to the datasets that include it
        - Create dictionary/standardised list of variables
            - Identify common variables first
            - Requires a research project to pull this together
            - Split workload up across clinical groups/studies and create agreed list

        - Define how to determine common outcomes from the variables
        - Find out what already happened and re-use/use as baselines – possibly other trials
        - Map similar items across similar datasets
- Ask data providers to create lists and share widely
    - HDR UK to coordinate on behalf of community
    - Dialogue between data providers/the 4 nations to harmonise
- Create virtual workshops to disseminate knowledge
    - Create cost neutral model/charge
    - Comprehensive understanding through NHRI learn to make more visible
- Provide clarity on what data isn't available
    - Identify experts on database linked into resource that can say if database will answer specific questions
    - BHF DSC bring this data together – expertise across datasets
    - Create Q&A system/needs filtering
- Approach speciality associations and national data providers
- Increase use of HDR UK Gateway
    - Make it easier/more streamlined – guidance

Additional points

- Streamline access process
    - Standardise process
- Role of BHF DSC as facilitator, not holder of data
- How might we ensure data is valid to answer research question?
    - Data dictionary – and if not here is highlighted
    - QC of accuracy of data at national level
    - Cross reference with other datasets
    - Tools and tips – help researchers review data available
    - Transparency in how data is checked
    - Quality of data (e.g. HES) – accuracy based on accuracy of coders. Invested because impacts payment
    - Various local and national steps to validation of data
    - BHF DSC – signposting, awareness raised

How might we make the necessary bureaucracy, obtaining consent, and ethics easier to navigate?

Why

- Facilitate researchers accessing the data for trials
    - More trials will use data in constructive ways
- Feasibility checks would provide reassurance trial can be done – to research team and funders
- It currently puts people/researchers off from the start
- Allows two-way visibility on what doing and why

How

- For planning – access to linked data possibly via BHF DSC

- Expertise on what variables are in datasets to help researchers what they can/should request
    - Someone to quickly check/run feasibility for researchers
- Links to data dictionaries – what data items are in each dataset
- Remove certain restrictions on wording e.g. consent – especially if consent to use data already in form
    - Develop consent process/wording for access to data – get PPIE, data provider etc buy-in/build on what already exists
    - Guidance on consent wording in various situations

How might we ensure that clear guidance is available to help trials teams navigate the process of gaining access to RCHD?

Why

- Standardise costs
- More replicable
- Saves time and easier to process
- Easier to process by whoever controls data
- Need information from the data provider to help design better study
- Less of a blocker to access
- Smarter trial design picking up practical what data sets are where

How

- Data passport for fast tracking through approvals
- Convince data providers to offer a "better" helpdesk service
    - Including time, cost
- Create Unified application process when applying for data
    - What/how?
- Make the most of existing expertise - use knowledge from successful studies
    - Share learning from less successful studies
    - Create resource of successful studies
    - Develop courses in cardiovascular
- Timeframes - need guidance at required time and throughout process
- Advocate and support clear guidance on timeframes - patient recruitment study
    - One standardised procedure for all datasets - at least within countries

Additional points

- How might we create geographic footprint on what available /what can do? map guidance
    - Focus on common data assets that already exist?
    - Make it clear if you actually need the data

How might we make it easier to obtain an accurate and consistent cost estimate?

Why

- Expectation is to pay a reasonable cost for what we get

- If this is supporting NHS, should it be an NHS cost, not just research
- Share costs
- Get value from data
- Get data that supports research
- Increase grant applications
- Increase transparency

How

- Need to know what costs are based on
- Develop range of costs/ballpark figures representing different data requests
    - Determine how to create cost model for prospective data
    - Create parameters e.g. ready-made, bespoke, amount of data, for retrospective
- Create “ready-made” datasets to meet common needs with fixed costs
    - Could be curated
    - Can NHS Digital share existing curated/oven ready datasets
- Align users for curating prospective datasets to share costs
- BHF DSC talk to NIHR
- Broad based survey to get an idea of what typical datasets are

Additional points

- How might we work in partnership with NHS Digital etc around cost transparency and agree what’s reasonable?
- Think about other future studies and how data can be reused – share wider?
- Is it cheaper to take whole dataset or cherry pick?
- How might we have a big conversation about this?
- How might we explore if it should be NHS support cost rather than research cost?
- How might we meet challenges of timescales from outset for grant applications aligning to quote timeline
- How long does a quote last? Ideally 1 year – should at least last time taken to complete process
- Is it consistent over e.g. 1-3 years

How might we make it easier to "sell / make the case" for Routine Collected Health Data to funders/on grant application?

Why

- Avoids duplication of collecting data for a trial that is already collected
- Funder understands costs of using and linking data
- Facilitates cost efficient research studies
- Deliver studies in quicker timeframes

How

- Show research team know how to use these data
    - Use in feasibility and work up of trial
- Run eCRF data collection alongside RCHD to show where savings can be made
    - Outcomes are valid from RCHD – funders, regulators and trialists have confidence
    - Run or find validation work

- Accept can't collect everything to run trial. Change culture of researchers.

How might we provide evidence that RCHD can successfully support clinical trials?

Why

- Range of opinions on how useful these data are – convince funders, regulators
- Reduce duplication
- Trials become more straightforward and quicker
- Show RCHD from trials accurate as may change clinical practice
- Worth funding trials using this data because answered question

How

- All 'stakeholders' - funders, trialists, regulators, patients - understand usefulness of data
    - How do we tell the 'right' people about quality of data
    - What are the concerns of different groups on use of RCHD e.g. trialist perceptions, patient perceptions
- Look at previous trials that have been successful
- Show it has worked in wider healthcare research
- Use trial data to go back and confirm RCHD
- Influence what RCHD is collected? Datasets and types of data within datasets
    - Influence/lobby
- Feedback to improve RCHD
- Streamline process on how researchers access data
- Show the limitations of the data and what it can be used for
- Complexity in requesting data for different users
- Show how the data have utility now
- Work with data providers to provide confidence in how data used

How might we make it easier to consistently derive outcomes from RCHD?

Why

- Increased value of data/
- Reuse anaylsis and e.g. change time
- Make useful comparison in the future when have more trials
- Make data submission easier

How

- Make it clear how outcome is defined/share code in public domain
    - Decide what the core outcomes are - 10-15 most important?
    - Process to make it living/up to date resource
- Create standard time points for comparisons etc
- Trial data and code needs to be made available to allow validity assessments/comparisons
- Create consensus around core principles
- Core outcome sets already exist for cardiac - (BenG) - comet?
- Pull together what been done already and get consensus
- Construct database

- Develop algorithms to determine outcomes from combining variables from RCHD

Additional points

- How might we allow variation where needed?
- How might we make available off-the shelf algorithms, with examples of where they had been used
- How might we use information in addition to outcomes to compare different studies

How might we ensure that the staff working on clinical trials have the skills necessary to use RCHD?

Why

- Help staff progression/career
- Faster process/access
- Decrease costs (?)

How

- Create an accredited course (in accessing data, rather than analysing specifically)
    - Identify all roles of people who access data (not just technical background)
    - Self service elements as people become proficient/accredited
    - Trusted environments CTUs (too restrictive?), safe havens etc
        - Set up more environments where people can use these skills and the skills will follow
        - Ensure safeguarding/governance with public oversite to enable them to become advocates
- Create working groups to share experiences and knowledge
    - Individual accreditation – workshop, webinars, NIHR building for clinical research – overlap/join up?
        - Needs to not be too hard, needs to ensure equity
    - MRC have a working group – how to connect into this?
    - Create a forum – people sign up to it
- Scale up usability of software
    - Can we join up on licence costs/make software more widely available. Have a conversation with NIHR
    - Clarity on software needed (e.g. SAS, R ?)
- Identify approved centres routinely using data for staff to look for roles
- Develop NIHR open data platform – centralised, validated to visualise data quickly

Additional points

- How might we create a super centre of federated data?

How might the BHF DSC best serve the community in this area of data-enabled trials?

Why

- Support scale
- Coordinate centrally to help more trials
- Bring cohort of trials together to help with quality of data

- Support aim of running leaner/streamlined
- One UK centre to liaise internationally (data sharing, access etc)

How

- Feasibility checks for planning trials
- Streamlined process for requesting data
- Signpost researchers in data available and how to access data
- Coordinate bringing new datasets or types of data collected
- Expert advice for trial teams
    - How do we have the correct person/people in post to do this?
- Advise (coordinate?) on lobbying various national bodies -harmonise data
- Support for surgery trials within wider CV field
- Engage with local clinical networks
- Training and courses on guidelines/accessing data
- BHF DSC feedback ideas like passport to other groups for longer term gains
- Pull together resources - data sets etc
- Provide advice to those who need it /want it
- Provide a Stamp of approval for study
- Be on board with application (time/costing)
- IRAS/NHIR? passport
- Make sure its scaleable
- Data custodians to recognise - have informal already
- But don't want more bureaucracy