British Heart Foundation Data Science Centre

Led by Health Data Research UK

Data-Enabled Trials Survey Report

JACKIE MACARTHUR, LYNN MORRICE, CATHIE SUDLOW, MATTHEW SYDES

JULY 2021

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

Data-enabled trials survey report

Contents

1. Executive summary

The use of routinely-collected healthcare data in randomised clinical trials offers the potential to deliver more efficient and cost-effective trials. However, it also presents challenges, with a very small proportion (~3%) of clinical trials[1] estimated to be using this data. 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 clinical trials.

To investigate the use of routinely-collected data in cardiovascular disease (CVD) clinical trials, we carried out a survey to identify trialists and CVD clinical trials using or planning to use routinely-collected data, and explore the challenges faced. The aim of the survey was to guide the BHF Data Science Centre in where to focus our efforts to best support the cardiovascular clinical trials community, recognising that the challenges are likely to be similar in other disease settings so the impact of our efforts may be generalizable beyond CVD.

The survey was emailed directly to 130 contacts working in UK clinical trial units, distributed to professional groups via newsletters, and provided to the wider community via Twitter. Of 56 survey responses, 35 were from people currently or previously involved in CVD clinical trials using routinely collected data. Their responses indicate that trialists use or aim to use routinely-collected data across clinical trial stages, most commonly in long-term follow-up. A variety of datasets are being or have been sought, including secondary care data, death data, CVD registries and audits, primary care data, and prescribing data.

The survey responses highlight clear barriers to the use of routinely collected data in cardiovascular clinical trials. Half of the respondents reported considerable challenges with ease and timeliness of data access, and with data availability. The majority of respondents cited the processes and bureaucracy (both real and perceived) involved in gaining access to data as the cause of challenges. These processes were described as complex and time consuming, with little guidance or information available. Difficulties arose from differences across the UK nations in the applications processes, available datasets and coding standards. Other challenging data issues included data inaccuracies, lack of detail about datasets or specific data items, specifying data requirements and defining disease outcomes. The barriers identified in this survey prevent or deter trialists from using linked health data in their trials, and may lead to them reverting to conventional approaches.

We are now carrying out further investigations to determine how we can overcome the barriers to data access and utility identified in this survey to deliver better, faster and more efficient trials[2]. The survey identified over 20 CVD clinical trials that are using or planning to use routinely-collected data, which we are further exploring with a view to identifying trials that we could potentially support as driver projects.

[1] Lensen et al., Access to routinely collected health data for clinical trials - review of successful data requests to UK registries. *Trials*. 2020;21(1):398.
[2] HDR UK One Institute Strategy 2019/2020 https://www.hdruk.ac.uk/wp-content/uploads/2019/04/HDR-UK-One-Institute-Strategy-compressed-1.pdf (accessed 22 April 2021)

2. Survey aims

- Identify key challenges associated with the use of routinely-collected data in CVD clinical trials
- Obtain additional information about the challenges to help us prioritise our efforts in addressing the challenges according to stakeholder needs
- Identify CVD clinical trials using/planning to use routinely collected data that we could support as driver projects

3. Survey methods

Distribution

- Survey emailed directly to ~150 contacts in clinical trials units (CTU). Some email addresses were generic for the CTU, so ~130 individuals targeted.
- Also distributed via mailing lists and newsletters to professional groups and Twitter
- Several reminders tweeted

Timeline

- Started distributing survey on 24th February 2021
- Survey closed on 16th March 2021

Analysis

- Descriptive statistics to summarise answers to categorical questions
- Free text responses were processed by:
    - Splitting each free text response into phrases
    - Grouping the phrases into categories

4. Survey responses

- 56 responses
- 34 from direct email to CTU contacts, 16 from mailing lists/newsletters, 6 from Twitter

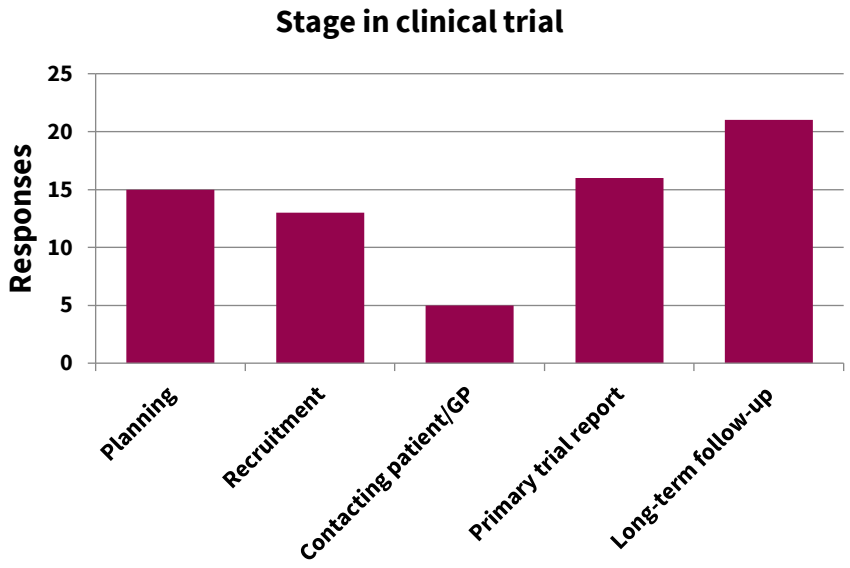
5. Survey results

Q1 - Have you been or are you currently involved in cardiovascular randomised clinical trials that have used/are using/plan to use routinely collected data?

- 56 responses
- 35 are currently involved in (29) or have previously been involved in (9) cardiovascular randomised clinical trials that have used/are using/plan to use routinely collected data
- 21 answered no and therefore were asked no further questions

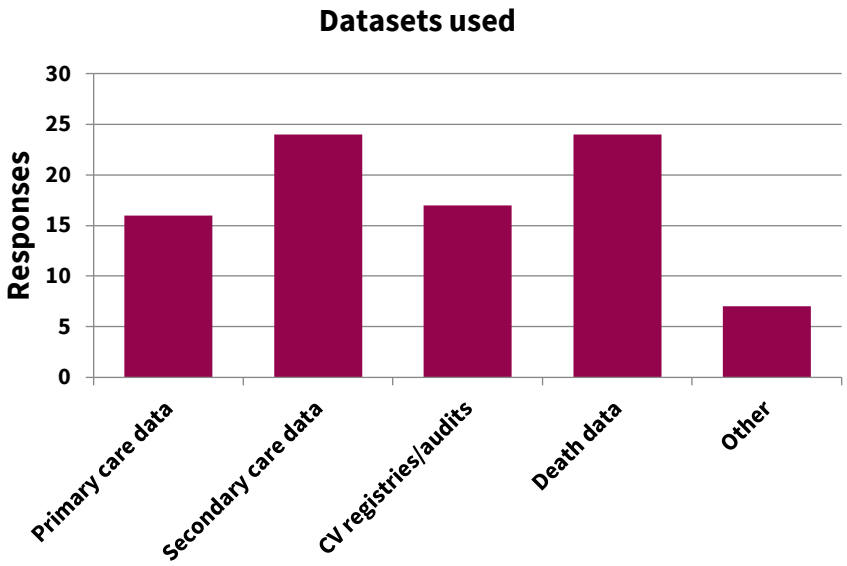
Q2 - What stage(s) in these clinical trials was/is routinely collected data used? (please select all stages that apply)

- 27 responses to this question



Q3 - What routinely collected datasets have been/would be of value to you in these trials? (please select all that apply)

- 27 responses to this question



- 7 “Other” responses
- 2 of these could have been categorised as “CV registries/audits”
  - NICOR
  - Registries of cardiac procedures
- 5 additional, of which 4 mention prescribing data and 2 imaging data
  - Community pharmacy drug dispensing data. National imaging data
  - Imaging
  - Prescribing, lab data, cancer registry, socioeconomic
  - National Prescribing Data (e.g. INTO Pharmacy)
  - Prescribing Data, Blood Transfusion Data, Haematology and Biochemistry Data

Q4 - Please provide additional details on how you used/plan to use routinely collected data in these trials

15 responses to this question

The statements below represent a summary of the range of uses described. They do not represent the full detail of the free-text provided by respondents.

- Plan trial design especially sample size calculations
- Identify eligible patients
- Recruit patient based on routine characteristics (primary care data)
- Primary and secondary outcomes of the study e.g. overall survival
- Verify outcomes of interest e.g. MI, HF admission, stroke and complications of interventions: bleeding, devices, infection; corroboration of participant-reported outcome events, identification of outcome events, monitoring prescribing intervention adherence, collection of baseline variables for adjustment of analyses.
- Follow-up patients for outcome assessment, long-term follow-up (secondary care)
- Baseline data and safety reporting (NICOR)
- Demographics, medications
- Assisting in locating trial participants lost to follow-up

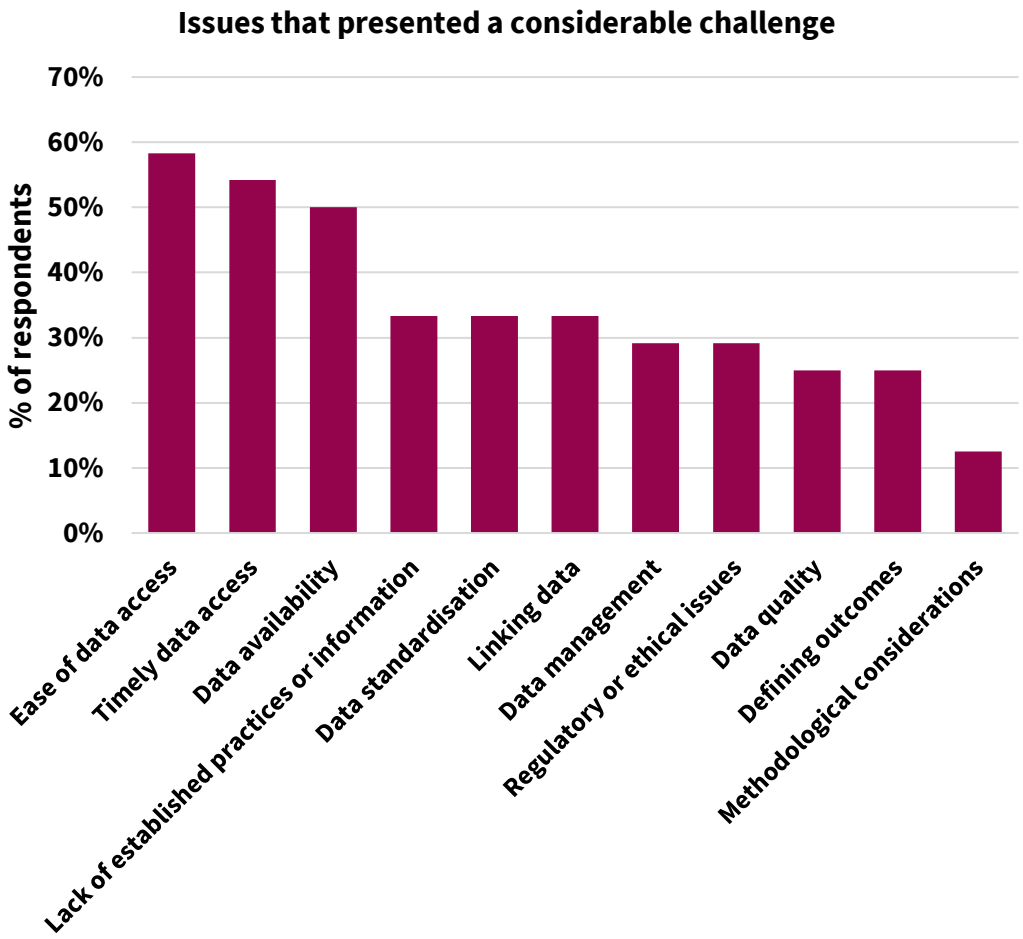
Selected full-text answers

- “Validation of RCD as an outcome measure in this patient population, versus adjudicated outcomes, to justify the use of RCD in RCTs in the future”
- “Long term observational study of patients undergoing cardiac surgery. Routine data is used for mortality tracing, baseline characteristics/demography, long term monitoring of outcomes (HES - admissions etc) and operation data (from Trust sources such as PHD/PATS)”
- “In our current TARGET-CTCA trial we are screening for eligible participants using a real-time data extract across six hospitals that enables a centralised pool of research nurses to recruit using proportionate consent on the telephone from a single site. We also use routine data sources for some outcomes (prescribing, hospitalisation and death) in Scotland, but have had to revert to conventional methods when expanding to sites in England as we try to overcome delays incurred by COVID-19.”
- “In previous trials (High-STEACS and HiSTORIC) all patients were enrolled by usual care clinicians using a form embedded in routine clinical care) and routine data was used for all characterisation of the participants, to support the adjudication of endpoints (console access to anonymised clinical letters and results from secondary and primary care), and for follow up (national and regional data across boards). As randomisation was at the institutional level in both trials individual patient consent was waived, and data from more than 80,000 consecutive patients was collected and reviewed.”

Q5 - What issues have you encountered/are you encountering in using or planning to use routinely collected data? (please rate each of these depending on how much of a challenge this issue has caused)

24 responses to this question

- As each issue was rated according to the extent of the challenge caused, there are different ways of plotting/analysing this data
- The same 3 issues dominate irrespective of whether issues are ranked according to the weighted average or the % of respondents that indicated that the issue presented a considerable challenge (plotted below)

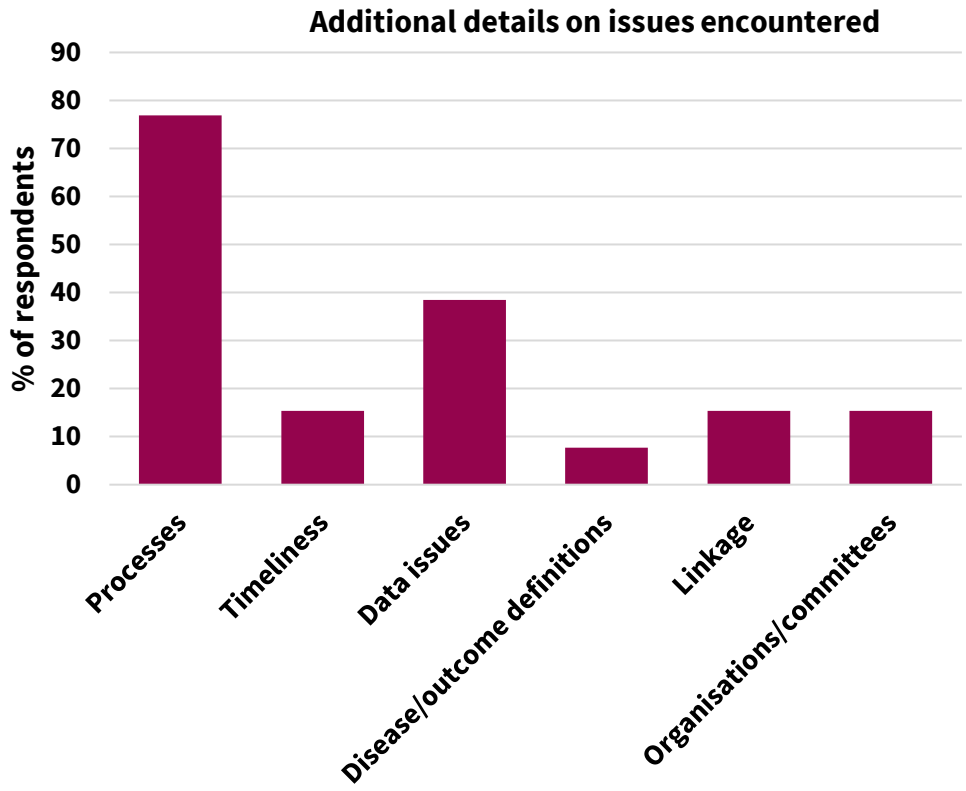


- 2 “Other” issues were specified, with the following detail provided
  - Governance and access across Trusts/Boards for unconsented data is challenging
  - Lack of real time flagging of individuals (considerable)

Q6 - Please provide additional details on any issues that you have found difficult to overcome or have prevented you from using routinely collected data

15 responses to this question

- The free text responses to this question were analysed by:
  - Splitting each free text response into phrases
  - Classifying each phrase according to the category of issue. Example statements under each category of issue are provided below.



Points below are example statements under each category of issue

Processes

- Getting data from NHS Digital was a nightmare. The processes are not facilitating research or the public good.
- Barriers to access i.e. despite consent was told this may not be valid (with change to NHS Digital processes) and that consent may not be specific enough.
- Not easy to understand or navigate systems across UK. Little guidance.
- Variable local interpretation of GDPR and information governance requirements.
- Questions regarding GDPR seem to go round and round.
- Approvals process excessively complex.

Timeliness

- The delays were enormous.
- The process to request the data initially is incredibly time consuming.

Data issues

- Data inaccurate or not detailed enough.
- Defining the data required is also very challenging, particularly without a comprehensive data dictionary. Provided with lists of field names, but not what these actually mean. When we get the data it's not always what we've asked for.
- Differences across UK nations in applications process, data storage and management, available datasets and coding standards – requires significant time etc to combine into single dataset.
- No central mechanism for obtaining haematology and biochem data.

Disease/outcome definitions

- Regulators concerned about the definition of disease and outcome.

Linkage

- Lack of information/clarity on linkage – “I don't even understand what DARS mean by linking data. I understand the concept of linking data, just not what they mean by it.”

Organisations/committees

- Too many organisations.
- Lack of continuity in NHS organisations.
- Lack of understanding by access committees.

Q7 - What would help you overcome these issues?

19 responses to this question

- Streamlined regulatory/applications processes
- Routine access to routinely collected data with ethical approval
- Clear information and guidance that covers all countries/regions. Guidance on application process.
- Access to data via universal database. Central data linkage hub.
- Easier linkages to outcomes data.
- Standardisation of ICD codes for outcomes. Standardisation on EHR definitions for diseases/outcomes.
- Data standardisation
- Better UK coverage
- Guidance for best practice

Q8 - Is there anything else you would like to tell us about your experience of using routinely collected data in randomised clinical trials?

11 responses to this question

- Routinely collected data are a fantastic resource, that can provide generalisable insights into care. If the system worked it could be massively beneficial.
- However access is frustrating, slow and expensive. You need to put work into curating the data and understanding it.
- Lack of clear guidance to trialists about the process, costs, approvals, and timelines across different regions that is a deterrent that will encourage trialists to revert to the conventional approach of using research nurses to manually collect data from each site.
- Training of clinicians and researchers is key.
- Why do we have so many codes for the same thing. Sometime less is more!
- We must fix this! Thanks for doing something about it. I've heard NHS data called the unicorn dataset - no one believes it exists

Q9 - What cardiovascular trials are you/have you been involved in that used/are using/plan to use routinely collected data?

14 responses to this question

- Over 20 CVD clinical trials that are using or plan to use routinely collected data were identified from the responses to this question
- Specific trial information is not being shared here due to some of these trials not yet having been approved and the risk of potentially identifying survey respondents from this information

Q10 - Which of the following categories best describe your current position? (please select all categories that apply)

21 responses to this question