
Peer reviewed version

License (if available): CC BY

Link to published version (if available): 10.1016/j.ejvs.2024.05.010

Link to publication record in Explore Bristol Research

PDF-document
Challenges of using routinely-collected healthcare system data in randomised trials

Christopher P. Twine,1,2 Haroon Ahmed,3 Fiona V. Lugg-Widger,3 Cherry-Ann Waldron,3 Matt J. Bown,4 Matthew R. Sydes5,6

1. North Bristol NHS Trust and University of Bristol, Bristol, UK
2. Division of Population Medicine, Cardiff University, Cardiff, UK.
3. Centre for Trials Research, Cardiff University, Cardiff, UK.
4. Department of Cardiovascular Sciences and the NIHR Leicester Biomedical Research Centre, University of Leicester, Leicester, UK.
5. MRC Clinical Trials Unit at UCL, Institute of Clinical Trial and Methodology, University College London, London, UK.
6. BHF Data Science Centre, Health Data Research UK, London, UK.

Corresponding author: Christopher Twine. Department of Vascular Surgery, North Bristol NHS Trust, Bristol, UK BS10 5NB. Email: Christopher.twine@bristol.ac.uk

Category of submission: Research letter.

Conflict of interest: None.

Word count: 811

Keywords: Clinical trials, randomized; analyses, cohort; clinical research methodology; diseases, peripheral arterial.

There is increasing interest in using individually linked, routinely-collected healthcare systems data for Randomised Controlled Trials (RCTs). These “data-enabled” trials aim to
reduce site and participant burden and increase efficiency of trial delivery. Routinely-collected data have been used successfully in large ‘big data’ retrospective studies for years, and new methodology such as target trial emulation has allowed researchers to reduce the bias inherent to retrospective data. This still cannot replace the traditional randomised trial design, but a hybrid approach potentially builds on the strengths of both designs.

During a recent grant application for a RCT of antithrombotic treatments following intervention for chronic limb threatening ischaemia, a major UK funding body suggested using routinely-collected data for outcome measure ascertainment. The suggestion was that this could reduce costs by reducing the need for individual patient follow-up using the clinical research network infrastructure in the UK. This is one of the driving premises behind data-enabled trials.

The proposed structure was fairly typical of surgical trials with a sample size of around 1,000, a follow-up time of 3 years and a composite outcome measure (acute limb ischaemia, major amputation, myocardial infarction, ischaemic stroke, and all-cause mortality) as a primary outcome. The composite was defined by stakeholders including vascular surgeons, interventional radiologists and patients. Such composites are increasingly common in trials, and funders often insist on this method of trial design. Simple outcome measures are becoming less common as a result.

All of these outcomes should theoretically be well recorded somewhere within hospital systems, national healthcare systems and hospital and general practice records. Outcome ascertainment using healthcare systems data was therefore explored in detail. We encountered various challenges that made a data-enabled approach to outcome
ascertainment non-viable within the limits of a traditional RCT funding application at this point in time.

First, RCTs in the UK need to be able to recruit from all four UK nations for equity. This requires at least four data access applications and approvals. There is inconsistency of data capture across the four nations and different information technology systems are used all adding time and cost.

Second, routinely-collected ischaemic limb outcome data validity and completeness is unclear. A recent study showed validation for basic ischemic limb data such as minor amputation in NHS England hospital admission data was poor with a low agreement with National Vascular Registry data where outcomes are inputted by vascular surgeons and interventional radiologists. Laterality for limb data can also be an issue. Apart from our primary composite outcome this would affect other common composites using ischaemic limb events in cardiovascular trials such as Major Adverse Limb Events (MALE) and occasionally Major Adverse Cardiovascular Events (MACE) depending on the definition used. Little is known about the quality or completeness of ischaemic limb outcome data in General Practice records.

Thirdly, lag for data availability (from data being inputted in a clinical environment to made available for use from a research database) can be substantial. As an example, the Clinical Practice Research Datalink, which contains data from UK general practices had relevant data available 9 months prior to the current date. This means that data required for the last day of the trial would only be available 9 months later. Data lag for linked secondary care data is up to 2 years. Such a delay adds significant cost to a trial, which may aim to recruit and follow-up within 3-5 years. The delay may also engender a delay to detecting important safety or effectiveness signals. Using routinely-collected data for this trial
application would have increased the explicit cost by around 16% (~£420,000) over using the ‘standard’ UK clinical research network infrastructure. There would have been a reduction in the unseen time and burden required to support clinical trials on site staff; the cost of which cannot be measured or clearly costed.

The sum of these challenges was that a data-enabled approach to outcome ascertainment for this particular trial would not have save time, resources, or costs, compared to the standard approach. The recent COMORANT-UK Delphi study echoed these challenges from a broader stakeholder base,\(^6\) implying they do not appear limited to surgical trials using ischaemic limb outcomes. More broadly, a lack of standardised outcome measures across trials (for example MACE is defined differently in almost every RCT), and coding between healthcare systems compounds the problem. There are few examples of successful data linked RCTs in the literature which highlights an international problem with data-enabled designs.

Work is required to move to a place where data-enabled trials can be performed more widely. Groups such as Health Data Research UK are seeking to change the infrastructure in collaboration with major funders. Key challenges include demonstrable data integrity and provenance and data retention, all consistent with the clinical trials regulations, and clear knowledge about data availability, timelines and costs. Funding bodies should usefully embed evaluative work to support moving to a data-enabled trials future within currently funded trials.

References