BMJ 2015;351:h4601 doi: 10.1136/bmj.h4601

EDITORIALS

Liberating the data from clinical trials

Liberated trial data have enduring potential to benefit patients, prevent harm, and correct misleading research

David Henry *professor*[1], Tiffany Fitzpatrick *meta-data specialist*[2]

[1]Dalla Lana School of Public Health, Institute for Health Policy Management and Evaluation, University of Toronto, and Senior Scientist, Institute for Clinical Evaluative Sciences, Toronto, ON, Canada; [2]Ontario Strategy for Patient-Oriented Research (SPOR) Support Unit, Toronto, ON, Canada

Despite the importance of reproducibility in research, clinical trials are rarely subject to independent reanalysis. In a linked paper, Le Noury and colleagues (doi:10.1136/bmj.h4320) have restored and reanalysed the controversial "study 329," which incorrectly portrayed paroxetine as an effective and acceptably safe treatment for children and adolescents with major depression.[1 2] The accompanying article by Doshi (doi:10.1136/bmj.h4629) details the miss-steps of the investigators, staff from the sponsoring drug company, the lead author's home academic institution, and the publication journal.[3] Study 329 is a model example for the movement to restore invisible and abandoned trials (RIAT), which calls on investigators to publish unreported trials and republish and correct misleading reports.[4]

In a recent review, Ebrahim and colleagues identified just 37 published reanalyses of clinical trials.[5] Only five were conducted by investigators not associated with the original report. A third of the reanalyses led to interpretations that were different from those of the original articles. In a recent blog, Ben Goldacre, co-founder of the +AllTrials initiative, which calls for all trials to be registered and published,[6-8] highlighted the example of an influential trial of intestinal "deworming" treatment. Reanalysis uncovered important errors and changed some central conclusions of the original report.[9 10] Goldacre applauds the original authors for having the courage to share their data, despite the potential for errors to be identified and having to go through the discomfort of their mistakes being made public.

While rare among clinical trialists, the idea of sharing scientific data is not new and is common practice within some disciplines, such as genomics, astronomy, and particle physics.[11] In a bold move by the standards of the time, researchers from the landmark Diabetes Control and Complications Trial (DCCT) made their data available to other investigators after they published the results of the original trial. To date, there have been over 220 ancillary studies using DCCT data, several of which have had an impact on the clinical management of diabetes.[12 13] These highlight the substantial added value that can be derived from sharing of trial data.

The move to access original trial data is part of the broadening open data movement in health, which has received support from major research funding agencies in the United States, Canada, the United Kingdom, Australia, and Europe.[11 14] Notably, the National Institutes of Health (NIH), which strongly encourage NIH funded investigators to share their data, provide secure data repositories for both clinical data and biological samples.[15] Recent reports from the Institute of Medicine (US), the Wellcome Trust (UK), and the Council of Canadian Academies argue for, and recommend, best practices to ensure safe sharing of clinical data.[16-18] And of course many journals, such as *The BMJ*, now encourage authors to make datasets available on request.[19]

Data sharing, however, is not without its risks.[18] As Ebrahim and colleagues point out, threats to patient confidentiality, data dredging with a risk of chance findings, and "rogue reanalyses" by investigators with their own agenda must be considered.[5] Data sharing also increases the responsibilities and burdens placed on investigators and institutions, for whom trials can become consuming, long term commitments. As illustrated by Le Noury and colleagues,[1] trial restoration can be a major undertaking for investigators carrying out the reanalysis, requiring substantial human and analytical resources.

Should restoration end with reanalysis, or should we do more? Data storage in repositories will enable independent researchers to repurpose trial data for new research questions—as shown by the successes of the DCCT.[13] If participants' data are stored with the identifying information needed to link to data stored in administrative claims or electronic medical record (EMR) databases, this will allow independent researchers to reactivate some "dormant" trials. Adding extra years of follow-up, via the linked databases, will allow the study of long term outcomes, including those not part of the original protocol.

The use of administrative and EMR data to capture clinical trial outcomes is becoming commonplace. Sometimes this has been part of the original trial protocol.[20-22] Less commonly, data linkage is used subsequently to capture additional years of follow-up. An important example of the latter is the 20 year

Correspondence to: D Henry david.henry@ices.on.ca

report of the West of Scotland Coronary Prevention Study (WOSCOPS).[23 24]

Reactivation of dormant trials will not be without barriers. In the case of older trials, data might have been destroyed, misplaced, exist in paper form only, or lack the variables necessary for linkage. The original trial consent forms might not have included permission to link the data, which will require research ethics boards to consider approval of "post hoc" linkage. In the case of the Canadian National Breast Cancer Screening Study, the research ethics board judged that the original approval to retrieve death certificates could be extended to electronic linkage to different outcomes by staff at Statistics Canada.[25]

The first step to reactivating dormant trials will be to identify which trials have been conducted and where the data are held. This can be done by searching trial registries and approaching research ethics boards, funding bodies, and investigators. Such an exercise has commenced in Ontario as a "meta-data mapping project," under the Strategy for Patient-Oriented Research (SPOR) initiative of the Canadian Institutes of Health Research.[26]

To enable reactivation of important clinical trials we will need to review some policies and procedures. Trial consent processes should routinely request permission to link the data to study long term outcomes. Stored data from participants should always include linkable fields, particularly health insurance numbers. Data management and retention policies should be reviewed to enable preservation of the data needed to enable long term follow-up of important clinical outcomes.

Most clinical trials are extremely expensive, and we believe that the pay-off from a systematic effort to reactivate selected clinical trials will be high and will further justify the original huge investments of time and money.

Competing interests: We have read and understood the BMJ Group policy on declaration of interests and declare the following interests: both authors work with the Ontario SPOR SUPPORT Unit, funded by the Canadian Institutes for Health Research (CIHR) and the Ontario Ministries of Health and Long-Term Care (MOHLTC) and Research and Innovation (MRI). Senior staff at the Ontario SPOR SUPPORT unit, the CIHR, MOHLTC, and MRI had no part in the writing of the article and have not provided any comments on it.

Provenance and peer review: Commissioned; not externally peer reviewed.

1 Le Noury J, Nardo J, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320.
2 Keller M, Ryan N, Strober M, et al. Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial. *J Am Acad Child Adolesc Psychiatry* 2001;40:762-72.
3 Doshi P. No correction, no retraction, no apology, no comment: paroxetine trial reanalysis raises questions about institutional responsibility. *BMJ* 2015;351:h4629.
4 Doshi P, Dickersin K, Healy D, Vedula S, Jefferson T. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ* 2013;346:f2865.
5 Ebrahim S, Sohani Z, Montoya L, et al. Reanalyses of randomized clinical trial data. *JAMA* 2014;312:1024-32.
6 +AllTrials. About AllTrials. www.alltrials.net/find-out-more/about-alltrials/.
7 Chalmers I, Glasziou P, Godlee F. All trials must be registered and the results published. *BMJ* 2013;346:f105.
8 Goldacre B. Scientists are hoarding data and it's ruining medical research. www.buzzfeed.com/bengoldacre/deworming-trials.
9 Davey C, Aiken A, Hayes R, Hargreaves J. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial. *Int J Epidemiol* 2015, July 22 (epub ahead of print).
10 Miguel E, Kremer M. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 2004;72:159-217.
11 Government of Canada, Policies and Guidelines. Open access—research data. www.science.gc.ca/default.asp?lang=en&n=2BBD98C5-1.
12 Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993;329:977-86.
13 Epidemiology of Diabetes Interventions and Complications (EDIC). EDIC publications. https://edic.bsc.gwu.edu/web/edic/publications.
14 Open Knowledge. What is open? https://okfn.org/opendata/.
15 National Institutes of Health. NIH data sharing repositories. www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html.
16 Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. National Academies Press, 2015.
17 Varnai P, Rentel M, Simmonds P, et al. Assessing the research potential of access to clinical trial data. Technopolis Group, 2015.
18 Council of Canadian Academies. Accessing health and health-related data in Canada, the expert panel on timely access to health and social data for health research and health system innovation. Council of Canadian Academies, 2015.
19 Loder E, Groves T. The BMJ requires data sharing on request for all trials. *BMJ* 2015;350:h2373.
20 Tu J, Donovan L, Lee D, et al Effectiveness of public report cards for improving the quality of cardiac care: the EFFECT study: a randomized trial. *JAMA* 2009;302:2330-7.
21 Steventon A, Bardsley M, Billings J, et al. Effect of telehealth on use of secondary care and mortality: findings from the whole system demonstrator cluster randomised trial. *BMJ* 2012;344:e3874.
22 Roberts M, Stokes J, King M, et al. Outcomes of a randomized controlled trial of a clinical pharmacy intervention in 52 nursing homes. *Br J Clin Pharmacol* 2001;51:257-65.
23 Packard C, Ford I, Murray H, McCowan C. Lifetime clinical and economic benefits of statin-based LDL lowering in the 20-year followup of the West of Scotland Coronary Prevention Study [abstract]. American Heart Association's Scientific Sessions. Chicago, 2014.
24 Barry SE, Dinnett E, Kean S, Gaw A, Ford I. Are routinely collected NHS administrative records suitable for endpoint identification in clinical trials? Evidence from the West of Scotland Coronary Prevention Study. *PLoS One* 2013;8:e75379.
25 Miller A, Wall C, Baine C, et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ* 2014;348:g366.
26 Ontario SPOR Support Unit. Work we support—data platforms & services. 2015. http://ossu.ca/work-we-support/data-platforms-services/.

Cite this as: *BMJ* 2015;351:h4601