Essays

Data sharing – peeping inside Pandora's box*

Moira A Hudson
Moirahudson7@gmail.com

DOI: 10.20316/ESE.2016.42.021

Abstract: In response to the growing consensus among funders, research foundations and government agencies to require the sharing of clinical trial data, the International Committee of Medical Journal Editors (ICMJE) has issued, and invited comments on, a proposal outlining it's requirements to support this obligation. In essence, for a clinical trial report to be considered for publication in an ICMJE member journal, authors will be required to share individual patient data within 6 months of publication. While on the surface this appears to be a clear-cut and fairly straightforward requirement, consideration of the implications reveals a plethora of issues, some of them potentially far reaching. In this article I explore some of the advantages and potential disadvantages of such a policy.

As laid out in the EASE Statement on Data Sharing, issued on 4 April 2016 "EASE is in agreement with the recently proposed requirement from the International Committee of Medical Journal Editors (ICMJE) that makes sharing of clinical trial data mandatory for manuscript acceptance by its member journals"[1, 2].

The benefits of data sharing and transparency are many. Analyses can be repeated, either to be verified, or in some cases, to be shown to be erroneous before they become completely embedded in the scientific record or have an effect on clinical practice; data can be re-used in addressing new but related research questions, thus reducing research waste[3] and falsification should become ever more difficult to get away with. All in all, data sharing can positively influence many things – retaining and possibly elevating the public's trust in science, the advancement of medical care, and the generation of new insights into clinical practice. Importantly, the benefits that the few grant to the many – results derived from clinical trials made possible by the patients who take part in them (often in the knowledge that they may be putting themselves at risk) are acknowledged. As an absolute minimum we, as a society, should ensure that full and judicious use is made of data obtained, and any potential benefit to the health of the populace is exploited.

Data transparency is not, per se, an entirely new thing. The European Medicines Agency made data sharing a legal requirement in October 2014 when it published its final policy on publication of clinical data[4], and this year we will see publication of the first compliant reports thus completing the journey from trial inception to completion with final results, down to the patient level, being completely transparent.

So, like many new initiatives, data sharing is a good one, and is to be applauded. But like all good ideas, the more one investigates it, the more complex it becomes. Even though I'm all in favour of transparency, for all the reasons given above, I can't help but think the concept may become so complex with all its various ramifications, as to be virtually non-implementable.

As I became more involved with the EASE Statement on Data Sharing, many questions presented themselves, and like others I came to realise that this isn't a simple matter of putting data in a database somewhere for all to access at will – there are many, many more issues to be considered, and potentially dozens of procedures to put in place. In fact, the whole process starts to 'grow like Topsy'[5].

> Clinical trials are far from perfect, but they remain, for now, the only feasible route we have to approval of life-changing/saving drugs.
> Do we really want to make this process any more difficult than it already is?

To "Protect and Serve" – do we need Data Police?

Recently, both before and after the ICMJE declared its intentions, claims were made by the medical publishing community that the reporting of clinical trial data is nowhere near as consistent as it ought to be. It's a requirement for publication of results by the vast majority of journals (that adhere to the ICMJE recommendations) that a trial is registered, but how many registered trials never actually report their results? How many patients potentially put themselves at risk in such trials? Kent Anderson's[6] recent offering to the Scholarly asks the question "Why is ClinicalTrials.gov still struggling?" and goes on to say that a large proportion of clinical trials registered may fail to report results. Anderson's sample was taken from the years 2007–2010 and is US-centric, but one could suppose that a similar trend would prevail in Europe and in other parts of the world. Looking at the most recent data in ClinicalTrials.gov[7] it is clear that the number of registered trials with posted results increases over time, as one would naturally expect, but it's impossible to work out what proportion of trials this represents – and indeed what proportion of trials never publish data. The EudraCT statistics 2016 (PDF)[8] makes no mention of reporting of results.

Shortly after Anderson published "Why is ClinicalTrials.gov still struggling?" the MNT Knowledge Center produced an adapted media release[9] in which they describe a study conducted by two medical physicists on radiotherapy trials. Perez-Alija and Gallego found that of 802 trials with

a primary completion date before 1 January 2013, 655 or 82% had not published even a summary result[10]. When they looked at trials that began prior to the passing of the FDA Amendments Act (FDAAA) that, as well as registration, mandates the deposition of additional information in the ClinicalTrials.gov database (including expanded information on the trial protocol and information on the results[10], the picture was hardly different with 76% failing to deposit results. Reasons for not publishing summary results may exist and may be valid, for example a trial may have been granted an extension, but this is not reported.

This leaves us with an uncomfortable truth – even in the face of legislation a large proportion of trials are failing to report results, and although there are small studies in particular therapy or disease areas, we really don't have the answer to the bigger question – how many trials never publish results?

Possibly the closest we can get to a real-world assessment of this sorry state of affairs is AllTrials[11] – All trials registered. All results reported – that claims that around half of all clinical trials ever conducted never report their results. Fronted by and the initiative of the maverick Ben Goldacre, AllTrials is an international collaboration with many organisations including the *BMJ*, Centre for Evidence-based Medicine, Cochrane Collaboration, James Lind Initiative, PLOS and Sense About Science, and aims to address this deficit. As of 29 June 2016, the AllTrials petition had been signed by 88,233 individuals, and 665 organisations had joined the AllTrials campaign. The question now is "when can we expect to see a measurable improvement in reporting?"

> When can we expect to see a measurable improvement in reporting?

Besides campaigning, how can we encourage (or incentivise) trial owners to comply with our pleas? Do we use the "carrot", and if so what form will the carrot take, or do we use the "stick" (the same question applies). One suggestion for the latter that has been aired is for research bodies to withhold further funding from any grant holder who fails to either publish trial data, or deposit it on some sort of accessible platform. For academics funded by public bodies that's probably enough of an incentive, it's a threat with consequences: report your data or risk your funding drying up. However, it is a rare thing for clinical trials to be purely academic, most trials necessarily involve extensive collaboration and funding by the pharmaceutical industry. We are all well aware of the tendency of pharmas to invest more in publishing positive data than neutral data, as it serves their purposes to do so, but what of negative data? That's important too, for many reasons. So long as a drug is shown to do no harm (ie the trial shows lack of efficacy and lack of toxicity) where's the incentive to invest to share in a timely manner? The issue of sanctions, financial or otherwise, now presents itself.

> The culture of reporting needs to be changed – but how?

Incentives (or punishments) aside, how is the reporting of trial data to be monitored and policed? Who will keep watch on the databases and trial registries? And who will dictate how long is "too long" before publishing data? When this deadline has been passed, what then? Will clinical trials that have no published data be posted on some "wall of shame"? I don't think anyone yet has the answer to these questions.

Which brings us to the issue of data re-use. Once trial data have been reported, who controls what then happens with those data? Do the authors or trial owners still have the last word on who can access and make use of their results? Will they have any say in what their data are used for? If the data are used, will their contribution be acknowledged?

The availability of individual-level patient data represents a huge potential bonus to health economists. Instead of informing their economic models with data generated by data simulation from the results of trials identified though systematic literature reviews, they will be able to obtain the original data, perhaps making their network meta-analyses and economic models more relevant to real-world clinical practice.

A less welcome potential development was explored by Longo and Drazen, in their editorial in the *New England Journal of Medicine* – the controversial issue of the emergence of "research parasites". These are individuals or groups who reuse data with the intention of pre-empting the research productivity of, or disproving the conclusions of the original investigators[12]. Trotter, a self-identified "data parasite" points that legitimate reuse of data is a necessity, and a good thing, and that it's happening on a wide scale in some areas (eg health economics) already[13].

So, while data re-use is to be applauded, not everyone is going to be happy with it, and it is important that safeguards be put in place to determine who may and who may not have access to a particular dataset – how can we ensure that the data won't be use inappropriately? An example would be covert use by medical insurance providers, perhaps to raise premiums on a certain demographic. In an ideal world this should, of course, not be an issue, but in this dog-eat-dog world where everyone wants a big profit gain for the smallest possible investment, the temptation to procure clinical trial datasets for exactly this purpose could be enormous.

Who foots the bill?

Like the expensive meal out, the bill for all this good intention has the potential to be the proverbial hot potato – everyone wants to eat, but no-one wants to offer up their credit card[14].

For a moment let us assume, that somehow and in an ideal word, all the results of every trial get reported either in the scientific literature as a manuscript, or in a database.

Who is going to fund these activities that in themselves have little potential to generate any revenue? Someone has to create the repositories, someone else has to load the data, another person (or persons) has to prepare the data (anonymising individual patient data as necessary) before it even gets to the point where it can be deposited. One can reasonably assume that any trials now in progress are unlikely to have collected data in a form that would make this easy – although that is expected to change over time. In the case of industry-funded trials, the costs of all this storing and cataloguing would be met by the pharma – does this mean a consequent increase in the price of drugs, and an added burden to public healthcare systems? Pity the poor academic PI and his team (or charity funded research teams) who has to devote precious research hours and funding to comply with these extra demands. To steal a quote from Kent Anderson's piece[6]:

> "When I was in grad school, I had to write a paper and publish it. Now, people are suggesting that I also pre-register my experiments; curate and upload all my raw data (which may be in non-standard or proprietary formats); deposit pre-prints; publish the actual paper in a peer-reviewed journal (because that's not going away); promote it through social media; upload it into sites like Academia.edu or ResearchGate; update my publication information in databases like ORCID, ImpactStory, and institutional measures; and watch for comments on post-publication peer review sites like PubPeer and engage with them as necessary".

To any rational human being that sounds like a big ask, doesn't it? And it leads as do most economic arguments to a division of a society into the "haves" and the "have nots": those who have resources available will comply, while those who lack such resources will struggle.

By the time we have considered the extra costs of all the storing, cataloguing, registering, and gatekeeping of the data, and the inevitable monitoring of the databases (how else will we know which trials have yet to comply with the data sharing initiatives?) will the potential gains of sharing data and therefore reducing waste be exceeded, ie will be we in "benefit negative equity"?

The patients' voice

Let's not forget the patients, without whom no trials would be possible. To date, the only mention of the patients in this debate, so far as I can tell, has been that their participation in clinical trials will be justly served by having the data published. When patients sign up for a clinical trial they do so on the understanding that the results will be used for the benefit of the common good, because quite often they themselves stand to gain little – but would they feel the same knowing that their data – their individual data – were available to scrutiny? I suspect that many wouldn't care less, but in some cases I can imagine that patients may not want their clinical details to be identifiable. To use the same example as before, an obvious issue would be access to information on heritable diseases by insurance companies. In such cases is it possible to guarantee anonymity? It all boils down to an issue of trust – can we trust those who get access to patient-level data to treat it with the respect it deserves. In most cases, probably yes, but it's that word "most" that presents the uncertainty.

Disclosures

The views expressed in this article are the author's own and do not necessarily reflect those of EASE.

The author is a member of EASE Council and was involved in writing the EASE Statement on Data Sharing.

References

1 EASE Statement on Dara Sharing. Available at: http://www.ease.org.uk/wp-content/uploads/data-sharing-statement.pdf (Accessed 20 September 2016)

2 Taichaman *et al*. Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors. *Annals of Internal Medicine*. 2016. doi:10.7326/M15-2928. Avaliable at: http://www.icmje.org/news-and-editorials/M15-2928-PAP.pdf (Accessed 20 September 2016)

3 *BMJ* blog. Paul Glasziou and Iain Chalmers: Is 85% of health research really "wasted"? http://blogs.bmj.com/bmj/2016/01/14/paul-glasziou-and-iain-chalmers-is-85-of-health-research-really-wasted/. (Accessed 20 September 2016)

4 The European Medicines Agency (EMA). Clinical data publication. 2016. Available at: http://www.ema.europa.eu/ema/?curl=pages/special_topics/general/general_content_000555.jsp. (Accessed 20 September 2016)

5 Uncle Tom's Cabin (1851-52) by Harriet Beecher Stowe

6 Anderson K. Why Is ClinicalTrials.gov Still Struggling? Available at: https://scholarlykitchen.sspnet.org/2016/03/15/why-is-clinicaltrials-gov-still-struggling/. (Accessed 20 September 2016)

7 Clinical Trials.gov. Trends, Charts, and Maps. Available at: https://clinicaltrials.gov/ct2/resources/trends. (Accessed 20 September 2016)

8 EudraCT statistics. Available at: https://eudract.ema.europa.eu/statistics.html. (Accessed 20 September 2016)

9 Medicalnewstoday. Massive failure to publish trial results exposes patients to risks without providing benefits for others. 2016. Available at: http://www.medicalnewstoday.com/releases/309684.php. (Accessed 20 September 2016)

10 Food and Drug Administration. Amendments Act of 2007. Public Law No 110-85, 2007

11 AllTrials. Available at: http://www.alltrials.net/.

12 Longo DL, Drazen JM. Data Sharing. N Engl J Med 2016; 374:276-277; DOI: 10.1056/NEJMe1516564.

13 Trotter F. Perspective: From a Data Parasite. 2016. Available at: https://tincture.io/perspective-from-a-data-parasite-5e96dc66ebcd#.ih3jvmrdr. (Accessed 20 September 2016)

14 The Free Dictionary. Definition of hot potato. Available at: http://www.thefreedictionary.com/hot+potato. (Accessed 20 September 2016)

* Pandora's box: an artefact in Greek mythology from the myth of Pandora's creation in Hesiod's Works and Days.