The many uses of data in public clinical trial registries

Gayatri Saberwal

Registries were primarily set up to (i) increase the enrollment in trials and (ii) decrease the bias in the published literature. Additionally, scholars have used registry data to answer a wide range of questions, related to: (i) transparency and information dissemination; (ii) healthcare; (iii) science; (iv) biomedical innovation; (v) the ethics of the trial enterprise; (vi) fulfilling regulatory requirements; (vii) the economic implications of clinical studies; (viii) the globalization of the trial enterprise and (ix) holding biomedical journals to account. With the help of numerous examples, I have documented some of these actual, or occasionally potential, uses of registry data. We should be aware of these many uses since they highlight the importance of (a) establishing and maintaining registries, (b) ensuring that trial records are comprehensive and accurate, and (c) funding research on registry data, and facilitating such research in other ways.

Keywords: Clinical trial, ethics, innovation, registry, transparency.

The many uses of data in public clinical trial registries

PUBLIC clinical trial registries have been set up to host, and make available to the public, the many details of trials. Such details include the condition, intervention, name of the Principal Investigator (PI), name of the sponsor, locations of trial sites, expected enrollment, expected start and stop dates, and so on. Trials may be registered prospectively, that is before the enrollment of the first patient, or retrospectively, that is after enrollment has begun.

Calls for the establishment of registries began in the 1970s (ref. 1). One call was made with the view that registration would facilitate increased patient enrollment. The second was to address the problem that studies with positive outcomes tend to be published more than others. If a trial was registered in a public repository at the start, then this practice could be monitored, or even stopped, thereby preventing a bias in the literature. ClinicalTrials.gov (CTG), the registry of the United States (US) government, and the ISRCTN registry were established in 2000 (ref. 2), and since then many others have been set up3. Seventeen registries, that fulfill certain criteria of the World Health Organization (WHO), have been labelled ‘primary registries’.

Registries do enhance enrollment in trials and decrease bias in the published literature. Additionally, scholars have used registry data to answer a wide range of questions, related to: (i) transparency and information dissemination; (ii) healthcare; (iii) science; (iv) biomedical innovation; (v) the ethics of the trial enterprise; (vi) fulfilling regulatory requirements; (vii) the economic implications of clinical studies; (viii) the globalization of the trial enterprise and (ix) holding biomedical journals to account. Below, we document some of these uses. Occasionally we highlight commentaries on the uses of registry data, or list potential uses. Unless otherwise specified, all studies relate to trials registered with CTG.

This article is primarily based on the academic biomedical literature. I have not considered (i) patient registries, that are concerned with specific medical conditions, or (ii) reports on clinical trials per se. Instead, I have focused on the above-mentioned nine categories of uses of registry data. I use the words ‘trial’ and ‘study’ interchangeably.

Transparency and information dissemination

A public registry’s primary purpose is to provide free access to the particulars of a large number of trials. This ensures that an application for study-related information does not have to be made to the sponsor or the regulator. Although most national, public registries are in English, the primary registries of Brazil, China, Iran, the Netherlands and South Korea4 are also in local languages. Such registries enable different audiences to access the hosted information. However, studies related to transparency have also been performed using registry data. Some examples are as follows.
A registry may contain contact information for the PI of a study. This information in CTG was used to contact academic PIs, to ascertain their views on the fields – proposed by WHO and the Ottawa Group – to be included in other registries.

Due to the highly structured nature of each registry record, it may contain more details than the subsequent publication. One study compared the reporting of methods in Clinical Trials Registry – India (CTRI) to that in publications, and found this to be so.

A commentary has pointed out that there is an increasing push to share Individual Patient Data, or IPD, through registries. This will enable researchers independent of the sponsor and regulator to analyse trial data and will improve the accountability of the enterprise overall.

An analysis has been carried out of the sponsors, conditions, interventions, and so on for almost 400 Expanded Access and Compassionate Use programs. Such analyses would help decision making by health charities, sponsors, and governments in particular.

Researchers have kept tabs on the number of trials running in a country, and tracked changes over time, or profiled the national trial landscape. Recently, the number and nature of studies related to Covid-19 have been closely monitored and several registries (such as the Australian New Zealand Clinical Trials Registry, CTG, and the European Union Clinical Trials Register) facilitate a search of their records for Covid-19 trials specifically.

Although we are not aware of such use thus far, registry data should help to understand trial disasters. To have to depend on the sponsor or the regulator for information is to risk the non-availability of information, or a significant delay in receiving it. The latter situation was exemplified by the case in which activists spent years trying to obtain information about the studies that underlay the approval of the drug Tamiflu to treat influenza.

In one study, researchers explored the stem cell trial landscape by studying the data in several registries. In principle, governments or others could use such data to dampen the hype that often accompanies the development of novel technologies.

A report of the United Kingdom (UK) House of Commons has pointed out that trial transparency promotes the public’s trust. A breakdown in this trust would provide space to fraudsters, as has been seen in the anti-vaxxers’ opposition to vaccination, and clinical research cannot be performed without the trust of the public.

Health

An interventional clinical trial is defined as ‘any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes’. As such, the primary purpose of registries relates to the trials’ implications for healthcare. Below, we list examples of the analysis of registry data around the issue of health.

Several studies have examined whether there is a correlation between the fraction of trials for a particular condition, and its contribution to a country’s disease burden. Although there will probably never be an exact correspondence, some countries achieve a much greater correlation than others. Countries with large discrepancies may wish to find ways to boost the number of studies in underserved medical conditions.

Many trial protocols in the US require that participants be proficient in English. Since this has implications for non-English speakers’ participation in trials, one study has quantified this language requirement in over 10,000 trials. Trialists need to be aware that such exclusion of ethnic or racial minorities could affect the understanding of differential drug effects in different populations.

Registries have been used to study the enrollment of special populations in trials. (i) Although military personnel have specific health needs, trialists have faced challenges in enrolling them. A study examined whether there were significant differences in the recruitment challenges for military and non-military participants. (ii) Another study looked at whether pregnant women were included in industry-sponsored drug trials. Since 40–80% of pregnant women need to take a drug at some point, it would be concerning if the drugs they are prescribed have not been suitably tested.

In the US, the Expanded Access program enables access to drugs that are not yet approved by the Food and Drug Administration (FDA). Researchers looked at when requests for such accelerated access were granted with respect to the date of FDA approval to determine whether the goal of this program – of making interventions available after their safety and efficacy have been established – had largely been met.

Studies have looked into the distribution of trials, both within a country and internationally, to assess how many patients were able to access experimental interventions for their possible health benefits. For the same reason, studies have also looked at the change in the fraction of enrollment from a country over time. One study mapped all the counties in the US that hosted the studies for a particular condition. The same study examined the change in recruitment within the US over three years. Another study investigated how many industry-sponsored international trials were suspended in Brazil after the completion of recruitment in other developing countries.

Science

As mentioned above, one of the primary reasons for establishing registries is to increase patient enrollment. Whereas such participation may improve the participants’
health, it is more likely to benefit future patients. As such, registry data can be examined from the point of view of science, as some studies have done.

If each planned study is registered, then it is easy to track the selective reporting of the results of those with positive outcomes. Indeed, the establishment of registries has decreased the bias in the literature. This leads to a more complete research base on which recommendations for medical practice can be made.

Prospective registration allows a comparison of what was planned in the trial and what was reported in a subsequent publication. Reconciling the two versions of the data helps to correct the published record. Studies have also examined whether trials were registered prospectively and how to increase prospective registration.

One study looked into whether trials were registered within 30 days of the beginning of enrollment. The greater the delay in registration, the greater is the potential to view the preliminary results and to alter the primary or secondary outcomes that will be measured.

A commentary on the data in CTRI stated that a given PI had been involved in 25 trials. There are concerns as to whether an overloaded PI can conduct ethical and scientific studies.

One study examined the failure to enroll patients in trials, as this could lead to underpowered studies.

Researchers have examined the nature of interventions for a particular condition. If industry tends to sponsor only one category of intervention, then charities or the government may wish to boost the number of trials of other, possibly more cost-effective, categories.

Registry data have also been used to evaluate the relative efficacy of various interventions.

Most trials in the US must make public their results within one year of collecting all data related to the primary outcome(s). Since the registry has a precise record of when a study ends, it helps to enforce the timely release of results, thereby helping to keep the published literature up-to-date.

A trial may need to be registered in multiple registries. Each registry allot it a unique number, and trialists are required to cross-reference the earlier registrations. However, this is not always done, leading to ‘hidden duplicates’. In a systematic review, such a study may be counted more than once, leading to a bias. Studies have attempted to quantify the problem of hidden duplicates.

Landscape analyses identify health conditions for which trials have not been undertaken, thereby identifying research gaps that need to be filled.

Randomized controlled trials (RCTs) are considered the gold standard of trial design. Are the trials running in a country RCTs, and has this changed over time? The government needs to be concerned about this.

With the increasing globalization of clinical research, there is concern that research findings from other populations may not extrapolate to Western patients. A study of the globalization of trials could help quantify this potential problem.

Although there is an academic effort to collate the various documents related to a given trial, it has been argued that it is better for science overall if the documents are available in stable repositories. Public registries are ideally suited to this task and should be reliably funded as a form of scientific infrastructure.

Innovation

Trial registries are a source of information on advanced biomedical innovations, as illustrated by some studies below.

Researchers have examined the trials funded by the US’s National Institutes of Health (NIH). Such an analysis lets NIH institutions know where they stand with respect to each other, in terms of the number and type of trials that they are piloting through the clinic.

In one study, researchers searched several registries for trials of novel stem cell therapies, novel treatments by known stem cell treatments and agents that help to stimulate stem cell action. They investigated which states in the US hosted such trials, whether industry was a funder or collaborator and the geo-temporal trends of these studies. Public health bodies in the UK quite often use registry data as part of their ‘horizon scanning’ to know which novel treatments may become available soon. Various stakeholders, such as funders of innovation, healthcare providers, scientists, doctors, policymakers and so on would use such information.

Researchers have examined the landscape of interventions for a particular condition, and have concluded whether or what needs to be done in terms of innovation.

Researchers have studied the effect of alliances on the entry of big pharma into new therapeutic areas.

Ethics

Good registry records help to maintain the ethical standards of the trial enterprise and various studies have examined registry data from this point of view.

One study investigated unapproved trials registered with CTG, which involved patients paying thousands of dollars to receive unapproved stem-cell treatments. Because they were listed on CTG, patients believed that these treatments were endorsed by the government. The CTG site now prominently lists a disclaimer that this is not so. This should reduce such fraud.

There are ethical consequences of being unable to enroll the planned number of participants, as many have already been enrolled, and will have to be discontinued in case the study is terminated. In one year alone, tens of thousands of patients were enrolled in trials that could not
answer the research questions because they ended prematurely or failed to achieve at least 85% of the planned enrollment\textsuperscript{11}.

One study was able to quantify the non-reporting of the results of large trials only because they were registered\textsuperscript{46}. To fail to report the results of a study, which may have involved thousands of patients, is a form of research misconduct. Such non-reporting also damages the trust between patients and both the investigators and the ethics committees\textsuperscript{1}.

Countries with less rigorous regulatory oversight have more vulnerable participants, and studies of the globalization of trials could quantify this issue\textsuperscript{59}.

**Regulatory issues**

Registries data have been used to help identify studies that are in breach of regulations.

A health advocacy group used CTRI to identify a trial run by a multinational company in India, that allegedly contravened the law\textsuperscript{50}. This led to the government ordering an investigation into the matter.

The above-mentioned phenomenon of hidden duplicates has been used to determine whether trials that are registered with CTG have not been registered with CTRI, although this is required by law\textsuperscript{57}.

Researchers have found that hardly any registry has the fields needed to describe the quality of an Investigational Medicinal Product, or IMP, used in post-marketing studies. Those that do, help to meet regulatory requirements\textsuperscript{51}.

One website\textsuperscript{40} uses registries to track trials whose results have not been reported promptly, as required by the law, and provides the government actionable information.

In one study, investigators looked into how many international studies were suspended in Brazil, after the completion of recruitment in other countries\textsuperscript{27}, and inferred that the slowness of the Brazilian regulatory process had deterred companies from completing their trials locally.

**Economic implications**

Aside from the benefits to science and health, clinical trials are an economic activity. The following examples relate to the economic implications of trials.

One study identified multiple unapproved stem-cell trials, listed on CTG, that required patients to pay thousands of dollars each, to receive unapproved treatments\textsuperscript{47}. Potentially, investigators could use this information to help quantify the scam.

There are economic consequences when studies are suspended, withdrawn, or terminated\textsuperscript{48}. These repercussions are felt not merely by study funders and collaborators, but also by countries\textsuperscript{27}. Analysis of the reasons for the failure to complete trials would help to address this challenge.

It is a waste of money if the results of a study cannot be used\textsuperscript{41}. Also, if a study is invisible because it was neither registered nor published, it may be repeated by another organization. As such, registries help to reduce research waste.

As mentioned earlier, one web resource\textsuperscript{40} uses registries to track those who have not reported study results promptly, as required by the law. The website also quantifies the economic loss to the government, from unimposed fines.

If industry only tests devices, for instance, for a given condition\textsuperscript{26}, then academic hospitals, charities, or the government may have to sponsor studies for other types of interventions. This has economic implications.

There are cost implications for any health budget if the interventions for a given condition are largely trialled by industry\textsuperscript{26}.

Pharma companies do not invest in products that are expected to have poor financial returns. Therefore, researchers have used registry data to model a variety of strategies to take forward a particular portfolio of interventions\textsuperscript{52}.

Studies of trials running in different countries have found wide divergence in the fraction of trials that involve industry as sponsor, funder or collaborator\textsuperscript{41,12,20}. Industrial involvement is a reflection of the commercial relevance of a trial, that has economic implications.

Which organizations are the major sponsors of trials\textsuperscript{53} and how has this changed over time? Governments should be interested in knowing this, and how their own companies are doing compared to those of other countries.

Studies have looked at the increase in the number of sites, the trial density, the share of trials in other parts of the world, and the changes over time\textsuperscript{39,54,55}. The globalization of trials has economic implications, and hundreds of millions of dollars have been saved per study conducted in Russia or India, for instance, instead of the US\textsuperscript{59}.

Registries document technologies in trials. Organizations planning their R&D value the competitive intelligence on others’ research programs. It is also helpful for public health bodies to know about innovations in trials as they plan their budgets\textsuperscript{53}.

Research findings in particular populations may not be extrapolatable to other countries\textsuperscript{39}, and if so, this puts in question the cost-effectiveness of off-shored studies. Although we are not aware of such usage thus far, registries ought to be helpful for any exercise to quantify this issue.

**Globalization and the developing world**

Although the various uses of registries mentioned above may be of interest to all countries, the developing world has specific concerns that registries help address. Examples of these are as follows.
In an international study, although it was planned that only 50% of recruitment would take place from India, in fact over 80% of the participants were from India\textsuperscript{32}. This renews concerns of exploitation of trial participants in developing countries.

The lack of local leadership from developing countries may cause the scientific and ethical integrity of an international study to be compromised. Therefore, researchers have analysed global trials to examine the extent of such leadership, as reflected in the authorship of publications\textsuperscript{56}.

Governments in developing countries wish to know whether sponsors prefer to conduct their trials in other countries, and studies have looked at the growth of trials in different regions of the world\textsuperscript{19,57}.

The governments of developing countries would wish to monitor the nature of trials taking place in their countries, especially when compared to those in developed nations. Researchers have examined the nature of studies – their size, duration and phase – conducted solely in high-income countries versus those conducted in other countries as well\textsuperscript{54}.

Holding medical journals to account

Journals are an essential part of the medical and scientific ecosystem, and editors play a crucial role in ensuring the quality of the literature. However there have been lapses in the performance of this gate-keeper function, and registries can be used to monitor some categories of errors. Examples of relevant studies are as follows.

In 2004, the International Committee of Medical Journal Editors (ICMJE) announced that for studies beginning 1 July 2005 onward, any trialist wishing to publish in their journals would have to prospectively register the study in a publicly accessible registry\textsuperscript{58}. However, journals did not always enforce this rule\textsuperscript{59}.

Even journals that endorse the Consolidated Standards of Reporting Trials (CONSORT), have not always enforced these standards. It was found that trial methodologies were reported better in CTRI than in subsequent publications, including those that appeared in journals that endorse CONSORT\textsuperscript{6}.

Another study looked into discrepancies between what outcomes were planned, as mentioned in a registry, and those that were described in published papers\textsuperscript{59}. The identification of such discrepancies helps to pressure journals to examine registry data, and compel trialists to present the planned analyses in their publications.

The same study also investigated how journals responded when discrepancies between registry and publication data were pointed out in real time, that is soon after the publications appeared\textsuperscript{59}. The finding that journals tend to respond slowly, if at all, also helps to pressure them to improve their practices.

In summary, researchers have posed an extraordinarily wide range of questions of registry data in the areas of transparency; health; science; innovation; ethics; regulatory issues; economic implications; globalization; and holding journals to account. We should be aware of these many uses as they highlight the importance of (a) establishing and maintaining registries, (b) ensuring that trial records are comprehensive and accurate, and (c) funding research on registry data, and facilitating such research in other ways. There have been many efforts on each of these counts, but much remains to be done.


22. Egleston, B. L. et al., Characteristics of clinical trials that require participants to be fluent in English. Clin. Trials, 2015, 12, 619–626.


29. Goldacre, B. et al., COM Pare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. Trials, 2019, 20, 118.


35. Tse, T., Fain, K. M. and Zarin, D. A., How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider. BMJ, 2018, 361, k1452.


37. Kumari, S., Mohan, A. and Saberwal, G., Hidden duplicates: 10s of 100s of Indian trials, registered with ClinicalTrials.gov, have not been registered in India, as required by law. PLoS ONE, 2016, 11, e0149416.


50. Mathieu, S., Comparison of registered and published primary outcomes in randomized controlled trials. JAMA, 2009, 302, 977.