



How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider

Tony Tse,¹ Kevin M Fain,¹ Deborah A Zarin¹

¹National Library of Medicine, National Institutes of Health, Department of Health and Human Services, Bethesda MD 20894, USA

Correspondence to: K M Fain
kevin.fain@nih.gov

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2018;361:k1452
<http://dx.doi.org/10.1136/bmj.k1452>

Accepted: 5 March 2018

ClinicalTrials.gov, a repository of information about clinical studies and their results, together with specialised search tools, provides a unique window into the clinical research enterprise, which includes all initiated, ongoing, and completed or terminated clinical studies. Researchers are increasingly using information from the database to assess research reporting practices, or to characterise the clinical research enterprise. Conducting valid analyses requires an understanding of both the capabilities and limitations of the database (that is, intrinsic factors) as well as reporting policies and other factors external to the database that influence the types of studies in ClinicalTrials.gov in a specified time. This article discusses 10 key issues that researchers need to consider when using the database to conduct research.

ClinicalTrials.gov is a web based resource providing access to summary information on publicly and privately supported clinical studies on a wide range of diseases and conditions. The database, which is maintained by the National Library of Medicine at the National Institutes of Health (NIH), consists of a clinical study registry and results database—two principal components of the trial reporting system.¹ The goal of the trial reporting system is to provide a

public mechanism for identifying and characterising all trials conducted to answer specific biomedical questions (that is, the so-called denominator for all such trials) and their summary findings (that is, the evidence base). Sponsors and investigators are responsible for submitting information about their studies to ClinicalTrials.gov. Ideally, complete and accurate key study information (eg, study design, recruitment information) is registered at the start of the study and updated throughout the research life cycle, and summary results are reported after the study is completed.² Achieving these goals depends on consistent and systematic reporting by sponsors and investigators. In practice, despite a substantial increase in reporting rates, continued suboptimal adherence to these principles means that accurate, complete, and timely information is not being provided for all studies.^{3–5} ClinicalTrials.gov currently contains registration information for nearly 270 000 studies in over 200 countries, and has posted summary results information for over 30 000 registered studies.

ClinicalTrials.gov is designed to provide a public listing of initiated, ongoing, and completed studies, and to serve as a source of summary results information to complement the medical literature. The original focus was on facilitating identification and retrieval of information about specific studies on investigational drug products for potential study participants. Over time, other benefits of reporting have become apparent including: fulfilment of ethical obligations to study volunteers who expect that participation will contribute to medical science; disclosure of all prespecified primary and secondary outcome measures, including those not reported in the peer reviewed publications, to demonstrate adherence to the study protocol; and promotion of more efficient allocation of research resources by identifying gaps and overlaps. It has also become clear that the database provides a unique resource for those studying the clinical research enterprise. For example, 404 research articles published between 2010 and 2015 reported using ClinicalTrials.gov data to study various topics, such as analyses of studies on specific conditions, and to support research on ethics and adverse event reporting.⁶

Proper analyses require an understanding of the database structure, its evolution over time, the organisation of study records, and the various incentives and requirements that shape its content. We are often asked to review analyses that used ClinicalTrials.gov information and have developed 10 issues to consider based on our experience from these consultations. For additional information, we provide links to resource material in table 1, and screenshots to help researchers use ClinicalTrials.gov in the online supplementary appendix.

SUMMARY POINTS

- ClinicalTrials.gov provides access to summary information on publicly and privately supported clinical studies on a wide range of diseases and conditions, and can be used to analyse research questions
- To use ClinicalTrials.gov as a data source for research, researchers need to understand the database structure, its evolution over time, the organisation of study records, and the various incentives and rules that shape its content
- Ten issues for researchers to consider are presented to help them use the ClinicalTrials.gov database more fully and consider the scientific appropriateness of their designs/methods to strengthen and expand public knowledge in important research areas.

Table 1 | Linked information for key ClinicalTrials.gov resources

Category and description	URL
Database overview	
ClinicalTrials.gov background	https://clinicaltrials.gov/ct2/about-site/background
How to use search results	https://clinicaltrials.gov/ct2/help/how-use-search-results
How to read a study record	https://clinicaltrials.gov/ct2/help/how-read-study
About the ClinicalTrials.gov results database	https://clinicaltrials.gov/ct2/about-site/results
How to find results of studies	https://clinicaltrials.gov/ct2/help/how-find/find-study-results
Trends and charts	https://clinicaltrials.gov/ct2/resources/trends
Terms and conditions for using ClinicalTrials.gov data	https://clinicaltrials.gov/ct2/about-site/terms-conditions#Use
Key policies	
Selected key policies	https://clinicaltrials.gov/ct2/about-site/history
FDAAA and the final rule requirements	https://clinicaltrials.gov/ct2/manage-recs/fdaaa
Final rule information	https://prinfo.clinicaltrials.gov/
NIH policy on the dissemination of NIH-funded clinical trial information	https://clinicaltrials.gov/ct2/manage-recs/resources#NIHPolicy
ICMJE trial registration policy	https://clinicaltrials.gov/ct2/manage-recs/resources#InternationalCommittee
Data element definitions and data entry templates	
Registration, expanded access, and results data element definition documents	https://clinicaltrials.gov/ct2/manage-recs/resources#DataElement
Registration template	https://prinfo.clinicaltrials.gov/Interventional_Study_Protocol_Registration_Template_Jan_2018.pdf
Simple results templates by module	https://clinicaltrials.gov/ct2/manage-recs/how-report#ScientificInformation
Quality control review criteria	
Registration and results quality control review criteria	https://clinicaltrials.gov/ct2/manage-recs/resources#ReviewCriteria
Data analysis tools	
Downloading XML and other data formats	https://clinicaltrials.gov/ct2/resources/download
Search API	https://clinicaltrials.gov/ct2/about-site/link-to#LinksThatSearchClinicalTrials
Research using ClinicalTrials.gov data	
Selected staff publications	https://clinicaltrials.gov/ct2/resources/pubs
API=application programming interface; FDAAA=Food and Drug Administration Amendments Act; ICMJE=International Committee of Medical Journal Editors; NIH=National Institutes of Health; XML=eXtensible Markup Language.	

History of ClinicalTrials.gov

The ClinicalTrials.gov study registry was launched in February 2000 in response to US federal law requiring the NIH to “establish, maintain, and operate a data bank of information on clinical trials for drugs for serious or life-threatening diseases and conditions . . . in a form that can be readily understood by members of the public”.⁷ The database later included other registration requirements such as the International Committee of Medical Journal Editors (ICMJE) policy (table 2). The US Food and Drug Administration Amendments Act of 2007 (FDAAA) further extended the scope and legal requirements of the ClinicalTrials.gov registry, and mandated the creation of a results database, which became operational in September 2008.¹⁰ Regulations for implementing the FDAAA, which were issued in September 2016 and became effective on 18 January 2017, clarified ambiguous statutory provisions, and expanded the scope and requirements of the results database considerably.^{11 13} To support the overall mission of the trial reporting system beyond its legal mandates, ClinicalTrials.gov permits and encourages registration and results’ reporting of all biomedical or health related research studies (and information on expanded access to experimental interventions) on humans that the sponsor or principal investigator indicates is in conformance with:

- Any applicable regulations on the protection of humans, or ethics review (eg, institutional review board approval), and
- Any other applicable regulations of the national or regional health authority.

In parallel, the number of international trial registries and reporting requirements has also grown (eg, the European Union Clinical Trial Registry and EU registration and results reporting regulations¹⁴). The World Health Organization’s International Clinical Trials Registry Platform (ICTRP) search portal allows users to look for studies in the 16 primary registries of the WHO registry network and ClinicalTrials.gov, from which it downloads registration information weekly. The WHO ICTRP search portal identifies single studies registered in two or more registries (that is, duplicate registrations) only when sponsors or investigators list unique registry identifiers on corresponding study records, but misses other duplicate registrations.¹⁵ Thus, searching the WHO ICTRP search portal might retrieve multiple records of the same study, which would affect various counts (eg, number of unique studies, or participants enrolled). As of March 2016, about two thirds of all trials available through the WHO ICTRP search portal were registered on ClinicalTrials.gov.⁴

ClinicalTrials.gov database—the basics

Who is responsible for submitting study information?

The study sponsor is responsible for submitting, updating, and verifying study information to ClinicalTrials.gov throughout the study life cycle (that is, “the responsible party”).¹⁶ In general, the holder of an FDA investigational new drug application, or investigational device exemption for a study is

Table 2 | Scope of key policies for clinical trial reporting to ClinicalTrials.gov*

Policy (effective date)	Scope/requirements	
	Registration	Results reporting
Food and Drugs Administration Modernization Act (21 November 1997; ClinicalTrials.gov registry available starting on 29 February 2000)	Clinical trials of investigational new drugs for serious or life threatening conditions and diseases	N/A
International Committee of Medical Journal Editors ^{8†} (1 July 2005 for newly initiated trials; 13 September 2005 for ongoing trials)	All clinical trials as a condition for consideration of publication of results	As of December 2017: expected for clinical trials subject to the results reporting requirements of relevant funding and regulatory agencies, and encouraged for all other interventional studies ⁹
Food and Drugs Administration Amendments Act (FDAAA) ¹⁰ (27 September 2007)	Non-phase 1 clinical trials of FDA regulated drug and biological products, and non-feasibility trials of FDA regulated device products	Non-phase 1 clinical trials of FDA approved/licensed drug and biological products, and non-feasibility trials of FDA cleared/approved device products
Final rule implementing FDAAA in Title 42, Part 11 of the Code of Federal Regulations (42 CFR Part 11) ¹¹ (18 January 2017)	Non-phase 1 clinical trials of FDA regulated drug and biological products, and non-feasibility trials of FDA regulated device products	Non-phase 1 clinical trials of FDA regulated drug and biological products, and non-feasibility trials of FDA regulated device products, whether or not FDA approved, cleared, or licensed
NIH policy on the dissemination of NIH funded clinical trial information ¹² (18 January 2017)	Clinical trials funded by the NIH	Clinical trials funded by the NIH

N/A=not applicable; NIH=National Institutes of Health.

*For a more complete listing, see table in Selected Trial Registration Laws and Policies (<https://clinicaltrials.gov/ct2/manage-recs/background#WhyRegister>).

†Requires registration at ClinicalTrials.gov, or one of the primary registries on the WHO International Clinical Trials Registry Platform (<http://www.who.int/ictrp/network/primary/en/>).

considered to be the sponsor. Otherwise, the sponsor is the entity or individual who initiates and has authority over the study. A sponsor may designate a principal investigator as the responsible party in certain circumstances. However, if at any time the designated principal investigator can no longer serve as the responsible party, accountability reverts to the sponsor.

When is the information submitted?

Studies are generally registered at the start of the study, then updated as the study is conducted. Once a study is completed, summary results can be entered. Although documents containing the protocol and statistical analysis plan are required to be submitted to ClinicalTrials.gov as portable document format (PDF) files at the time of results' reporting for trials that are subject to the final rule implementing the FDAAA, they can be uploaded for any registered study at any time. The sponsor or principal investigator can update, correct, add, or sometimes delete information from the study record through the Protocol Registration and Results System¹⁷ at any time. Earlier versions of study records are accessible through the "history of changes" link under "study details" on each record (figs S1 and S2). The "tabular view" displays both the current (last updated and posted) and original (initially registered) entries for the primary, secondary, and other outcome measures for easy comparison (fig S3).

How is the study information reviewed?

All study information, which is self reported by the responsible party through the Protocol Registration and Results System, must meet established quality control review criteria before posting.¹⁸ Automated validation rules prevent submission of records that fail to meet certain technical requirements. Quality control staff manually review the information for consistency with remaining quality control criteria (for details, see issue 5). A unique ClinicalTrials.gov identifier (NCT number) is assigned to each study record.

Ten issues to consider when planning an analysis

As with any scientific investigation that uses data in a resource designed to support different objectives, researchers must carefully consider if their research questions can be adequately answered using the information in ClinicalTrials.gov. Whether the investigation will describe practices in the trial reporting system (eg, how non-inferiority trials are reported¹⁹), or make inferences about the clinical research enterprise more broadly (eg, how study designs of cardiovascular trials differ from oncology trials²⁰), researchers need to consider important attributes of the database (eg, data element definitions, requirements), and related extrinsic factors (eg, prevailing incentives that affect the sample of studies that are included in the database at any point in time). In this section, we describe the 10 issues that researchers should consider when using ClinicalTrials.gov in research. The penultimate section of this article integrates some of these issues as dos and don'ts of using ClinicalTrials.gov data for research.

1. ClinicalTrials.gov includes more than "clinical trials"

Despite its name, the database includes three different types of study records.

- **Interventional (or clinical trial).** Participants are assigned prospectively to interventions as specified in a protocol to evaluate the effect of the interventions on biomedical and/or health related outcomes.¹⁶ US federal reporting requirements and the ICMJE policy apply to clinical trials. Currently, 80% of registered records and 93% of records with posted results are clinical trials.
- **Observational.** Biomedical and/or health outcomes are assessed in predefined groups of study participants receiving specific interventions, but not assigned by an investigator.¹⁶ Despite the lack of major reporting policies for observational studies, they account for 20% of registered records, and 7% of records with posted results.²¹
- **Expanded access.** US regulations require registration of expanded access to certain FDA regulated drug or

biological products for patients who do not qualify for enrolment in a clinical trial. ClinicalTrials.gov currently includes over 450 expanded access records.

Each study type is associated with a unique set of registration data elements (eg, the study phase data element is only available for interventional studies). For the remainder of this article, we focus on clinical trials. Table S7 shows the mandatory registration and results information for clinical trials by category as well as registration data elements that are part of the WHO ICTRP trial registration dataset. For a listing of all mandatory and optional data elements for all of the study types, see the ClinicalTrials.gov documents listing data element definitions.¹⁶

2. Follow-on studies may be registered as separate records

ClinicalTrials.gov defines a single study as a set of data collections and analyses governed by a single protocol that includes a plan to analyse the same group of participants; each study is registered separately and represented by a single record with a unique NCT number. In contrast, data collections or analyses that require re-consent and/or include participants who were not part of the original study are registered as separate studies. However, even though follow-on or extension study designs track participants after completion of an initial study to collect longer term data, responsible parties may register initial and follow-on studies as separate records. ClinicalTrials.gov currently does not provide a way to identify such pairs of records automatically, but clues can be found in certain data elements, including: title/acronym (eg, “ABSORB” and “ABSORB Extend”); brief summary (eg, “The primary objective is to assess the safety and tolerability . . . in participants . . . who completed Study 205MS301”); and eligibility criteria (eg, “patient successfully completed core study CRFB002H2301”).

3. Incentives for reporting trials change over time

Reporting incentives influence both the kinds of studies that are registered on ClinicalTrials.gov and the amount of information submitted, and change over time. Any sample of registered records, including those with posted results, reflects the prevailing incentives for disclosing trial information, including laws, policies, and scientific norms/practices (table 2). It is important to consider how these evolving incentives might affect an analysis. For example, the fact that more device studies were registered in 2016 than in 2004 probably reflects, in part, the effect of the ICMJE policy and FDAAA, which makes it difficult to determine if there had been an actual increase in the number of device studies conducted using information from ClinicalTrials.gov.

Similarly, the content of registered records also reflects the prevailing incentives. For instance, FDAAA required information about primary and secondary outcome measures, but did not specify timeframes for outcome measures. Timeframes became mandatory

data elements in ClinicalTrials.gov in December 2012. Without this incentive, before December 2012, responsible parties often did not provide information about timeframes even though ClinicalTrials.gov had offered the option.²²

4. ClinicalTrials.gov includes mandatory and optional data elements

Each record contains both mandatory and optional structured data elements. Mandatory data elements comprise the minimum amount of information needed to understand the study and its findings (table S7) and are annotated with red asterisks in the documents listing data element definitions.¹⁶ Automated validation rules prevent the submission of records with missing entries for data elements that were mandatory when registered, and identify obviously inconsistent information at the time of submission (eg, recruitment status set to “recruiting,” but listed study start date is in the future). Optional data elements are not reported in all posted records.

Data elements, subelements, and submission requirements have evolved (table S8). For example, the primary outcome measure data element, introduced as an optional free text field in 2005, was restructured in 2007 to include three separate optional subelements for greater precision that paralleled the top three levels of the specification framework for reporting outcome measures: title, including domain (eg, anxiety) and specific measurement (eg, Hamilton anxiety rating scale); description, including specific metric (eg, change from baseline); and timeframe for the time point(s) at which the measurement is assessed for the specific metric.²³ All three subelements became mandatory for study records submitted after 1 December 2012.

5. ClinicalTrials.gov conducts quality control review

Quality control review criteria ensure that entries are complete and meaningful, and show internal consistency and face validity. These criteria with examples are described in publicly available documents.¹⁸ Several approaches are used to ensure consistency and standard implementation of the quality control review criteria across the review staff, including a comprehensive training programme, standardised review comments, and regular auditing of reviews by review staff members. At registration, the most common problem identified by the quality control review is insufficient specificity of outcome measures (eg, “safety”). At results reporting, the five most important problems by frequency of occurrence are invalid/inconsistent unit of measure, insufficient information about a scale, internal inconsistencies between different parts of a record, inclusion of written results or conclusions, and unclear baseline, or outcome measure.²⁴

Because quality control review criteria evolve with policies, scientific practices (eg, new study designs), and experience, inconsistencies with the current criteria may exist in records posted before

implementation of the relevant criteria. For example, older records may include primary outcome measures that do not meet current requirements for specificity.

Researchers must also keep in mind that, as with the peer review process, quality control review cannot ensure the veracity of the submitted information, or determine whether the entry is fully compliant with various policy or legal requirements. For example, discrepancies and inconsistencies have been found between summary results posted on ClinicalTrials.gov and other sources of results for the same trials (eg, peer reviewed publications, FDA review documents).²⁵⁻²⁷ Although in many cases the data in ClinicalTrials.gov are more complete and more structured,²⁸ there is no way to determine which (if either) version of the results is correct.

6. Records can be modified by the responsible party at any time

Responsible parties can add, edit, and sometimes delete information from a record at any time, although all previous versions are publicly available through the archive site. While some data elements are expected to change regularly over the study life cycle (eg, recruitment status), others change infrequently (eg, study design). ClinicalTrials.gov permits changes to allow the information displayed in records to reflect the current state of a study, and to ensure accuracy. Thus, researchers need to extract and save all data elements from an analysis to ensure future access to the dataset (eg, for audit or reanalysis). They also need to state clearly in their manuscripts when the data were extracted for analysis. Despite the capability to make changes to a record, many records are not updated in a timely way. For example, as of February 2018, nearly 25 500 of 270 000 (9.4%) records list an “unknown” recruitment status, which indicates that posted information for a study previously listed as “recruiting,” “not yet recruiting,” or “active, not recruiting” has not been updated or verified within the past two years.

The most recent entry submitted for each data element is displayed on ClinicalTrials.gov, but all earlier versions of that record back to the initial registration can be reviewed using the “history of changes” archival function in “study details” (fig S1), or the “change history” feature in the “tabular view”. Depending on the goals of the analysis, it might be necessary to use information from historical versions of the record (eg, characterising the nature and frequency of main changes in ClinicalTrials.gov records after initial registration²⁹).

7. ClinicalTrials.gov does not have all information for all studies

An important overarching factor to consider is that ClinicalTrials.gov will always be incomplete in two ways: first, individual studies may be missing from the database and, second, study information may be missing from the records.³⁻⁵ Because reporting policies are not comprehensive, and compliance will never be

perfect, some trials in the clinical research enterprise are not registered with ClinicalTrials.gov, or do not have any posted summary results. Thus, researchers using this database need to consider the representativeness of any sample of records, the types of inferences that can be drawn about the clinical research enterprise (see issue 9) from an analysis of the sampled records, and possible biases.

Even when reported, posted records on ClinicalTrials.gov may contain incomplete information. One reason is that certain data elements may not have existed, may have had a different format or structure, or may have been considered optional at the time the study information was initially submitted (see issue 4). Even though ClinicalTrials.gov permits the modification of records at any time (see issue 6), incentives are often insufficient to motivate responsible parties to update entries in older records (see issue 3). Therefore, researchers need to be aware of when data elements of interest were introduced and became mandatory, and the existing incentives. For example, an investigation of the reporting of primary outcome measures over time would need to consider the timing of various changes to the data element and the associated requirements.

8. ClinicalTrials.gov data can be accessed in several ways

There are three ways to identify and retrieve a set of study records directly from ClinicalTrials.gov:

- Basic fielded search: specify term(s) by condition/disease, facility location, or other terms (fig S4).
- Advanced search: enter search terms in any of over 20 structured fields, including study type (eg, “interventional”, “observational”), recruitment status (eg, “recruiting”, “completed”), intervention/treatment, and has posted results (fig S5).
- Download the search results: registration and results information from study records can be downloaded in a spreadsheet format, or as a single zip file containing the study records formatted using XML that researchers can later search with their own system and tools (fig S6).

Alternatively, researchers may prefer to use the Clinical Trials Transformation Initiative’s database for aggregate analysis of ClinicalTrials.gov.³⁰ This database contains the full set of registration and results information from ClinicalTrials.gov; researchers should note the update schedule of the database to determine its appropriateness for their particular analysis.

9. Defining a sample of records to answer a specific question

Sample selection is integrally related to the analytical goals, and any intended inferences. Many analyses rely on ClinicalTrials.gov data to reach conclusions about the practices and characteristics of the clinical research enterprise, but the degree to which a sample of ClinicalTrials.gov records accurately represents the full population of studies within the clinical research enterprise varies by the prevailing reporting incentives (eg, by geographical region, over time, and the type

and characteristics of the study). For example, a recent press release announced that “Korea ranks 6th in world in number of clinical trials” based on information from ClinicalTrials.gov.³¹ The analysis apparently did not account for the fact that ClinicalTrials.gov does not include all trials (issue 7), incentives for reporting by geographical location vary and change over time (issue 3), or the existence of a primary registry in the WHO ICPT registry network “for clinical trials (researches) to be conducted in Korea”.

In contrast, an investigation of publication rates among NIH funded trials used a sample of ClinicalTrials.gov records for the analysis.³² Given that multiple reporting incentives apply to such trials (see issue 3), including FDAAA, ICMJE policy, and NIH’s recommendation that grantees register their trials, the evidence suggested that the sample from ClinicalTrials.gov was largely representative of NIH funded trials.

10. Using the ClinicalTrials.gov results database

Results information available from ClinicalTrials.gov is more complex than registration information. Some additional factors need to be considered when using the results database.

- Obtaining and using data. Unlike registry data elements, which can easily be downloaded in a list or spreadsheet format (records as rows and data elements as columns), most of the results data elements are only available in XML because of the high degree of variability in the data structure of results data tables (eg, numbers of rows and columns). However, full protocols, statistical analysis plans, and informed consent forms can be downloaded as PDF files. ClinicalTrials.gov started accepting these documents on 29 June 2017.
- Understanding the results reporting requirements. Researchers need to understand the results reporting requirements, which are different from the registration requirements, as well as changes to these requirements and relevant dates when the requirements became legally effective. For example, although the deadline for results information is generally one year after the primary completion date (date of final collection of data for the primary outcome), results submission may be legally delayed by up to two more years in certain circumstances,¹¹ and not required at all in other circumstances. Additionally, in some circumstances—for example, when data collection is ongoing for a secondary outcome measure at the primary completion date—partial results information for a study must be submitted until complete results information has been provided.
- Policies and compliance. When assessing the compliance of studies with results reporting policies, researchers should be clear about the relevant standards for the analysis, such as general ethical principles versus more specific legal requirements for reporting. Determination of compliance with FDAAA or the final rule requires detailed understanding of the law, the regulations, and their implementation.

Dos and don’ts of using ClinicalTrials.gov data for research

1. Don’t assume that studies registered on ClinicalTrials.gov are an unbiased reflection of the clinical research enterprise

- Studies reported to ClinicalTrials.gov are an incomplete sample of the clinical research enterprise. The degree of bias introduced by a sample of records varies according to the strength of the incentives and norms for reporting studies with certain characteristics. Key factors include intervention type (eg, drugs versus behavioural interventions), funding source, date of study initiation or completion, geographical location, and regulatory jurisdiction.
- Researchers conducting analyses aimed at showing differences or changes in the clinical research enterprise need to consider carefully how reliance on the samples of studies available from ClinicalTrials.gov might introduce biases in the analyses.

2. Don’t use the wrong data elements to operationally define a sample

- Many concepts used in analyses of the clinical research enterprise can be addressed by several different data elements in ClinicalTrials.gov. The choice of data elements will affect the meaning and the interpretation of an analysis—sometimes substantially.
- Before establishing an operational definition for a concept, researchers should fully review the definitions of the data elements, and understand the consequences of all possible options; examples of several common concepts are shown in table 3.

3. Think carefully when selecting an analysis population

- Many analyses can be summarised as ratios using the number of studies with a certain feature as the numerator, and the overall number of studies of interest (analysis population) as the denominator. The selection of an analysis population influences the validity of any conclusions and their relevance to the research question. For example, when determining the percentage of studies that have results posted on ClinicalTrials.gov, choosing all registered studies as the analysis population provides a methodologically straightforward finding but does not account for important factors. Selection of all studies completed since the start of the results database, or those legally required to report results to ClinicalTrials.gov as the analysis population (and comparable numerators) would estimate the usefulness of the results database in providing summary results and compliance with the law, respectively.
- The choice of analysis population is also important when using a data element that is missing from many studies (eg, a new or optional data element). A denominator of all studies versus a denominator of only those studies that have a value entered for that data element will provide different results.

Table 3 | Common concepts and implications of selecting different data elements for a concept

Concept and options	Comments
Study initiation	
Study start date (estimated or actual)	Before 18 January 2017, the format for study start date was month and year
Date on which overall recruitment status changed to "Recruiting"	Determining the date that recruitment status changed is technically challenging and requires the archive site; retrospectively registered studies may have completed recruitment when first submitted (eg, listed as "completed")
First submitted date	First submitted date may be before or after study start date
End of study	
Primary completion date*	Before 18 January 2017, the format for primary and study completion dates was month and year
Study completion date*	Consider differences in definition of primary completion date (final collection of data for the primary outcome) and study completion date (final collection of data for the primary and secondary outcome measures and adverse events)
Date on which overall recruitment changed (eg, to "Active, no longer recruiting," "Completed" and/or "Terminated")	Determining the date that recruitment status changed is technically challenging and requires the archive site
Geographical location	
Based on location of sponsor, principal investigator, study sites, funder	Sample will vary depending on which criteria are used; consider appropriateness of each option for the analytical question
Funder	
Based on sponsor only, collaborators, secondary IDs, ClinicalTrials.gov-defined funder type, combination of these	Sample will vary depending on which criteria are used; consider appropriateness of each option for the analytical question

*On 18 January 2017, the format of the ClinicalTrials.gov data elements for dates changed from month and year to month, day, and year.

4. Don't forget that information about a given trial in a record may have changed over time

- Study records change over time as the study itself is implemented and completed. Decisions about when to use the currently displayed entry for a data element versus a previously submitted entry (available through historical versions of records from the archive site) should be based on the goals of the analysis. For example, assessing concordance between prespecified and published primary outcome measures requires the entry originally registered at the start of the study rather than the most recent, updated entry on the public site.

Conclusion

The ClinicalTrials.gov database is a powerful tool for understanding various aspects of registration and results reporting, as well as the underlying characteristics and practices of the trial reporting system, and clinical research enterprise. However, it is easy to misuse the information inadvertently and draw invalid conclusions, especially because the database does not contain all clinical trials in the clinical research enterprise. As with any other analysis using an existing data resource designed to support different goals, researchers must carefully articulate the specific research question and determine the suitability of the database to answer the question. It is important that researchers understand the characteristics and nuances of the database, including the evolution of reporting incentives. By following these recommendations, researchers can use the ClinicalTrials.gov database more fully in scientifically appropriate ways, which in turn can strengthen and expand public knowledge of these important research areas.

We thank Rebecca J Williams for her critical review of the revised manuscript and for the many helpful suggestions. The views expressed in this article are those of the authors and do not necessarily reflect the positions of the NIH.

Contributors: TT, KF, and DAZ conceived the idea, drafted the manuscript, or critically revised the manuscript for important intellectual content, gave final approval of the version to be published,

and are accountable for all aspects of the work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted

Funding: The authors are supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine.

Competing interests: We have read and understood the BMJ Group policy on declaration of interests and declare the following interests: none.

Provenance and peer review: Commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- 1 Zarin DA, Tse T. Sharing individual participant data (IPD) within the context of the trial reporting system (TRS). *PLoS Med* 2016;13:e1001946. doi:10.1371/journal.pmed.1001946
- 2 Zarin DA, Williams RJ, Tse T, Ide NC. The role and importance of clinical trial registries and results databases. In: Gallin JI, Ognibene FP, Johnson LL, eds. *Principles and practice of clinical research*. 4th ed. Academic Press, 2017:111-25.
- 3 Huić M, Marušić M, Marušić A. Completeness and changes in registered data and reporting bias of randomized controlled trials in ICMJE journals after trial registration policy. *PLoS One* 2011;6:e25258. doi:10.1371/journal.pone.0025258
- 4 Zarin DA, Tse T, Williams RJ, Rajakannan T. Update on trial registration 11 years after the ICMJE policy was established. *N Engl J Med* 2017;376:383-91. doi:10.1056/NEJMs1601330
- 5 Anderson ML, Chiswell K, Peterson ED, Tasneem A, Topping J, Califf RM. Compliance with results reporting at ClinicalTrials.gov. *N Engl J Med* 2015;372:1031-9. doi:10.1056/NEJMs1409364
- 6 Fain K, Zarin DA, Williams RJ, Tse T, Rajakannan T. Impact of results reporting in ClinicalTrials.gov. 38th Annual Meeting of the Society for Clinical Trials. Liverpool, United Kingdom, 2017.
- 7 Food and Drug Administration Modernization Act of 1997. Public Law 105-115. Secondary Food and Drug Administration Modernization Act of 1997. Public Law 105-115. <https://www.gpo.gov/fdsys/pkg/PLAW-105publ115/content-detail.html>
- 8 De Angelis CD, Drazen JM, Frizelle FA, et al, International Committee of Medical Journal Editors. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *N Engl J Med* 2005;352:2436-8. doi:10.1056/NEJMe058127
- 9 International Committee of Medical Journal Editors. Clinical trials—registration. <http://icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html#one>.
- 10 Food and Drug Administration Amendments Act of 2007. Public Law 110-85. September 27, 2007. <https://www.gpo.gov/fdsys/pkg/PLAW-110publ85/content-detail.html> <https://www.gpo.gov/fdsys/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf>

- 11 National Institutes of Health, Department of Health and Human Services. Clinical trials registration and results information submission. *Fed Regist* 2016;81 FR 64981-65157. <https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission>
- 12 National Institutes of Health, Department of Health and Human Services. NIH policy on the dissemination of NIH-funded clinical trial information. *Fed Regist* 2016;81 FR 64922-64928. <https://www.federalregister.gov/documents/2016/09/21/2016-22379/ni-h-policy-on-the-dissemination-of-nih-funded-clinical-trial-information>
- 13 Zarin DA, Tse T, Williams RJ, Carr S. Trial reporting in ClinicalTrials.gov—the final rule. *N Engl J Med* 2016;375:1998-2004. doi:10.1056/NEJMs1611785
- 14 European Union. Regulation (EU) No 536/2014 of The European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. *Official Journal of the European Union*. L 158; 27 May 2014. https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf
- 15 van Valkenhoef G, Loane RF, Zarin DA. Previously unidentified duplicate registrations of clinical trials: an exploratory analysis of registry data worldwide. *Syst Rev* 2016;5:116. doi:10.1186/s13643-016-0283-8
- 16 NIH. U.S. National Library of Medicine. ClinicalTrials.gov. Data element definitions. <https://clinicaltrials.gov/ct2/manage-recs/resources#DataElement>
- 17 ClinicalTrials.gov PRS. Protocol Registration and Results System. PRS user's guide. 31 October 2017. <https://prsinfo.clinicaltrials.gov/prs-users-guide.html>
- 18 NIH. U.S. National Library of Medicine. ClinicalTrials.gov. Review criteria and other support materials. <https://clinicaltrials.gov/ct2/manage-recs/resources#ReviewCriteria>
- 19 Gopal AD, Desai NR, Tse T, Ross JS. Reporting of noninferiority trials in ClinicalTrials.gov and corresponding publications. *JAMA* 2015;313:1163-5. doi:10.1001/jama.2015.1697
- 20 Califf RM, Zarin DA, Kramer JM, Sherman RE, Aberle LH, Tasneem A. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. *JAMA* 2012;307:1838-47. doi:10.1001/jama.2012.3424
- 21 Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? *CMAJ* 2010;182:1638-42. doi:10.1503/cmaj.092252
- 22 Viergever RF, Karam G, Reis A, Gherzi D. The quality of registration of clinical trials: still a problem. *PLoS One* 2014;9:e84727. doi:10.1371/journal.pone.0084727
- 23 Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database—update and key issues. *N Engl J Med* 2011;364:852-60. doi:10.1056/NEJMsa1012065
- 24 Dobbins HD, Cox C, Tse T, Williams RJ, Zarin DA. Characterizing major issues in ClinicalTrials.gov results submissions. Poster presented at: Eighth International Congress on Peer Review and Scientific Publication; September 2017; Chicago, IL. https://prsinfo.clinicaltrials.gov/publications/Characterizing_Major_Issues-508-compliant.pdf
- 25 Becker JE, Krumholz HM, Ben-Josef G, Ross JS. Reporting of results in ClinicalTrials.gov and high-impact journals. *JAMA* 2014;311:1063-5. doi:10.1001/jama.2013.285634
- 26 Hartung DM, Zarin DA, Guise JM, McDonagh M, Paynter R, Helfand M. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. *Ann Intern Med* 2014;160:477-83. doi:10.7326/M13-0480
- 27 Schwartz LM, Woloshin S, Zheng E, Tse T, Zarin DA. ClinicalTrials.gov and Drugs@FDA: a comparison of results reporting for new drug approval trials. *Ann Intern Med* 2016;165:421-30. doi:10.7326/M15-2658
- 28 Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. *PLoS Med* 2013;10:e1001566, discussion e1001566. doi:10.1371/journal.pmed.1001566
- 29 Pranić S, Marušić A. Changes to registration elements and results in a cohort of ClinicalTrials.gov trials were not reflected in published articles. *J Clin Epidemiol* 2016;70:26-37. doi:10.1016/j.jclinepi.2015.07.007
- 30 Clinical Trials Transformation Initiative. Improving public access to aggregate content of ClinicalTrials.gov. <https://aact.ctti-clinicaltrials.org/>
- 31 Chu M. Korea ranks 6th in world in number of clinical trials. *Korea Biomedical Review*. 2018 Jan 15. <http://www.koreabiomed.com/news/articleView.html?idxno=2355>
- 32 Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ* 2012;344:d7292. doi:10.1136/bmj.d7292

Web appendix: Supplementary appendix