Precision and Negative Predictive Value of Links between ClinicalTrials.gov and PubMed

Vojtech Huser, MD, PhD, James J. Cimino, MD
Laboratory for Informatics Development, NIH Clinical Center, Bethesda, MD

Abstract

One of the goals of translational science is to shorten the time from discovery to clinical use. Clinical trial registries were established to increase transparency in completed and ongoing clinical trials, and they support linking trials with resulting publications. We set out to investigate precision and negative predictive value (NPV) of links between ClinicalTrials.gov (CT.gov) and PubMed. CT.gov has been established to increase transparency in clinical trials and the link to PubMed is crucial for supporting a number of important functions, including ascertaining publication bias. We drew a random sample of trials downloaded from CT.gov and performed manual review of retrieved publications. We characterize two types of links between trials and publications (NCT-link originating from MEDLINE and PMID-link originating from CT.gov).The link precision is different based on type (NCT-link: 100%; PMID-link: 63% to 96%). In trials with no linked publication, we were able to find publications 44% of the time (NPV=56%) by searching PubMed. This low NPV shows that there are potentially numerous publications that should have been formally linked with the trials. Our results indicate that existing trial registry and publisher policies may not be fully enforced. We suggest some automated methods for improving link quality.

Introduction

ClinicalTrials.gov (CT.gov) is a registry of clinical research trials originally established by the US National Library of Medicine in 1999 to allow investigators to publish information about their research studies on the Internet for recruitment purposes.[1] Since then, trial registry has become a requirement by regulatory agencies[2] and publishers[3] to increase transparency of clinical research, mitigate publication bias and to facilitate subject recruitment.

As with most research, there is an expectation that trials would be registered in a trial registry and peer-reviewed publications will result from the trials registered in CT.gov. A publication can be connected to a trial by two types of links. The first type of link occurs automatically when an article describing a clinical trial properly references the trial ID or "NCT number" assigned by the trial registry. We refer to this link as an *NCT-link*. The second type of link is manual and is initiated by the principal investigator (PI) when he or she updates the trial registry data. CT.gov provides a designated optional field called "result reference" where a PI fills in an PubMed article ID (PMID) of an article which describes results of the trial. We refer to this link as a *PMID-link*. Both types of links achieve a similar purpose to connect trials to most relevant publications. The reverse relationship is also important when a publication is explicitly linked to one or multiple trials.

This link between ClinicalTrials.gov records and MEDLINE records has potential value for a number of purposes, such as determining which studies in CT.gov have led to publications, which manuscripts submitted to publishers have met the requirement of prior registry in an acceptable clinical trial registry, and locating original data related to particular publications. In fact, a number of prior studies have made use of this connection for these purposes.[4,5] However, the link is not automatic and depends on the author providing the correct NCT or PMID reference. In the case of the NCT-link, there is a multi-step process in which the NCT reference is passed from the authors to the journals and then from the journals to MEDLINE. Since this process involves one or more manual steps, there is the potential for erroneous or even missing information. Furthermore, a connection between the publication and the trial does not necessarily imply that the publication actually reports the results of the study. To our knowledge, the accuracy and meaning of the trial-publication link has not been studied. The purpose of this paper is to examine these issues.

Resources

Clinical Trials.gov

CT.gov is a registry operated by the National Library of Medicine with more than 120 thousand studies, as of February 2012. It provides the ability to download data about all registered trials in two formats. The first format is a tab-delimited file with the twenty most important trial parameters. A second, more detailed format, is an XML file about each registered trial. Single trial retrieval can be easily combined with other data sources to perform

sophisticated computerized analyses of integrated research data. For example, to obtain complete details for the trial with identifier 00084383, the following URL retrieves the complete XML data for that trial:

http://clinicaltrials.gov/show/NCT00084383?resultsxml=true.

ClinicalTrials.gov provides mandatory and optional elements for each trial.[6] Important mandatory elements (some of which are required by US law) include: trial title, trial type, trial phase, central contact, trial eligibility criteria, start date, completion date, and trial sponsor. Optional elements include: detailed description, PI, collaborators, study acronym, reference publications, reasons for study termination, and biospecimen retention policy. See Figure 1 for an example.

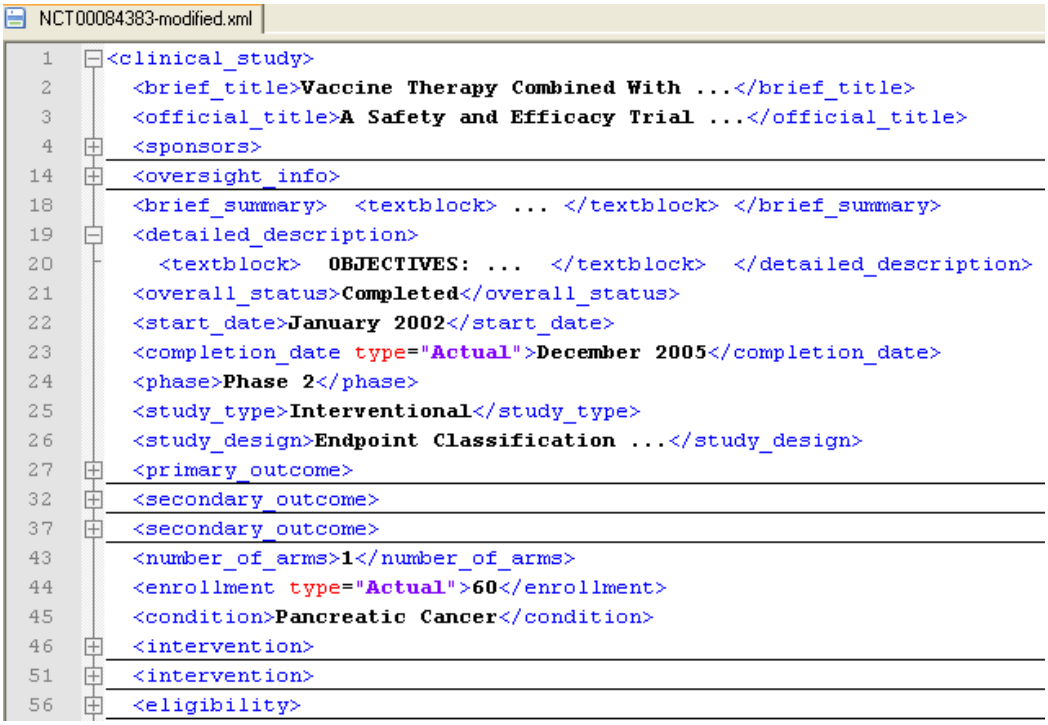

Figure 1. Example of the XML structure retrieved from ClinicalTrials.gov (modified)

Most importantly, each registered study is assigned a unique identifier (the "NCT number"), which is then used during recruitment, in publications that report results of the trial, or in other related publications (e.g., methodology descriptions and meta-analyses).

In an ideal case, the author registers the trial prior the trial start date (to benefit from CT.gov's ability to centrally advertise trials to potential participants and to comply with requirements of journal publishers for articles reporting results of clinical trials). The registry record is then typically updated at several time points by the trial sponsor or the PI. The first important milestone is when the study stops actively recruiting subjects and no longer should be advertised. Another milestone is when the trial collects all data for the primary outcome evaluation (CT.gov's "<primary completion date>" field), or when data for all evaluated outcomes were collected (CT.gov's "<completion date>" field). Based on those two dates, mandatory results reporting[7] of trial results may be required. Finally, the PI can optionally update the record with links to publications describing the trial's results (CT.gov's "<result_reference/PMID>" field).

PubMed

Since July 2005, the MEDLINE database captures secondary identifiers for each citation. This new citation attribute can be queried using the "[SI]" tag in MEDLINE and via the PubMed e-Utils data interface. The secondary identifier field contains accession numbers to various databases of molecular sequence data, gene expression, chemical compounds and clinical trials.

The SI entry is key for the first type of link (NCT-link). The trial identifier within the SI entry appears in PubMed within the Supplemental Information section as a direct link to CT.gov; e.g., "SI-ClinicalTrials.gov/NCTxxxxxxxx")[8] For example, the following URL can be used to search PubMed for all publications that reference NCT identifier NCT00360815:

http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=NCT00360815 [SI]

The second type of link (PMID-link) is also displayed in PubMed. Since both CT.gov and PubMed are provided by the National Library of Medicine (NLM), the PMID-link is properly rendered and up-to-date despite the fact that it depends on the update of data within CT.gov rather than MEDLINE. It is displayed only as "SI-ClinicalTrials.gov" but actually contains a link with a PMID-based search within CT.gov, and it is only displayed for publications where such a search indeed retrieves at least one linked clinical trial. See Figure 2 for an example.


Child Adolesc Psychiatry Ment Health. 2010 Feb 8;4:7.

Effectiveness of a single-session early psychological intervention for children after road traffic accidents: a randomised controlled trial.

Zehnder D, Meuli M, Landolt MA.

Department of Psychosomatics and Psychiatry, University Children's Hospital Zurich, Zurich, Switzerland. daniel.zehnder@kispi.uzh.ch

Abstract

BACKGROUND: Road traffic accidents (RTAs) are the leading health threat to children in Europe, resulting in 355,000 injuries

... ~ ...

LinkOut - more resources

Medical

ClinicalTrials.gov

Figure 2. PubMed view of article (PMID: 20181120) showing a LinkOut link to ClinicalTrials.gov

For example a MEDLINE article of PMID:20181120 (Figure 2), has the PMID-link URL:

http://clinicaltrials.gov/search/term=20181120[PUBMED-IDS]

However, for users of the PubMed eUtils data interface, it is important to note, that the link to the relevant trial shows only on the PubMed main web user interface and is not available in the XML data provided by PubMed eFetch service. Obtaining complete linkage data requires the user to merge data from PubMed and CT.gov.

Methods

Our study is motivated by the assumption that a researcher may start his or her review of literature by identifying all past relevant clinical trials within ClinicalTrials.gov and then search for published results of those trials. Our goal was to determine the quality of the links between the trial registry and published articles within MEDLINE.

We queried CT.gov to obtain all trial records with completion dates between September 1, 2004 and December 31, 2008. This criterion ensured that all analyzed trials had at least a three-year time window for their results to be published. Within this "CT.gov dataset", we identified all trials with linked publications using an R script that presented queries to the e-Utils data interface and CT.gov. The script also created a working spreadsheet file with pre-populated links to PubMed to facilitate manual review of any linked publications.

We further used two subsets of this dataset and describe a methodology to determine precision of links between CT.gov and MEDLINE and the negative predictive value (NPV) when no links are present.

Precision of links between Clinical Trials.gov and PubMed

We define *trial with linked publication* to denote trials for which one or more publications can be found in MEDLINE using the NCT-link or PMID-link described above (see Glossary section for overview of all trial terms we introduce).

For precision determination, we randomly selected 405 trials with linked publication from the CT.gov dataset. For each trial in this precision subset we retrieved the linked articles and we manually reviewed each article for relevance to the title and description of the trial from CT.gov. The relevance determination was initially based on the comparison of the article title to the trial title and trial parameters of investigated medical condition and interventions. Cases in which the title of the publication appeared to indicate that it was reporting on the results of the trial were immediately counted as true positive (TP) matches. In cases where the publication and trial titles did not appear to match, we examined the abstract and, where necessary, the full text of the article. In these cases, we noted whether the publication reported results of the trial (TP), was related to the trial but did not report the results (TP*) or was unrelated to the trial (false positive; FP). The manual review was conducted by two reviewers. We used a pilot study with a small set of trials to clarify the categories (TP, TP*, FP) and cases where the two reviewers disagreed were discussed and re-reviewed until agreement was reached. Informed by the pilot study results, the final manual review of 405 trials was divided among the two reviewers equally with no overlap. In was out of the scope of our study to measure inter-rater agreement.

Negative Predictive Value

We define *trial with no publications* to denote a trial for which there are no publications indexed in MEDLINE that report full or partial results of that trial using any search strategy or by contacting the trial principal investigator. Previous studies show that as many as 64% of trials are of this type.[5] We further define *trial with only not-linked publications* to denote a trial for which publications exist, but none of them are linked to the trial via the NCT-link or the PMID-link.

Our original intent was to complement the precision methodology by measuring recall of a link between trials and publications. However, it was not feasible to conduct large manual review of PubMed search results required for measuring recall. Instead, we chose to measure a negative predictive value of such link, since it required somewhat less extensive and a different type of review of PubMed search results. There is also a different gold standard dataset required for measuring recall versus NPV. Whereas recall requires identifying all relevant publications for a given trial, for NPV, the gold standard relies on manual verification that indeed no not-linked publications exist for a trial with zero linked publication.

NPV is defined as the number of trials that are true negatives (TN) divided by sum of the number of false negatives and the TNs, or TN/(TN+FN). We attempted to classify a random sample of 50 trials with no linked publication and non-empty PI field (sampling from the CT.gov dataset defined above) as *trials with no publications (TN)* and *trials with only not-linked publications (FN)*. Since we decided to start with trials with no linked publication, the trial category of *trial with both linked and not-linked publications* was not applicable. Including the mixed category would also require more complex comparison – namely of the overlap of results from linked-publication search and manual search.

Despite decreasing the manual review required by measuring NPV, we still had to make limiting assumptions for objective identification of any not-linked relevant publication for a given trial and provide a computational approach to facilitate such manual review. The NPV manual review was conducted by a single human reviewer. The ultimate gold standard for determining non-existence of relevant publications is contacting the trial PI. From prior studies we know that the response rate is too low. For example, Ross et al. contacted 117 PIs and received responses from only 44 (37.6%)[5].

We formulated three search strategies that attempted to retrieve relevant publications based on key trial attributes such as trial title, primary PI (last name and first initial), PI-provided trial keywords, and keywords derived from the trial-investigated medical condition and (for interventional trials) the trial interventions. The first search strategy (S1) used the PI and a date range with limit dates of trial start date and trial completion date plus five years. The second strategy had all S1 conditions ANDed with PI provided keywords connected by logical OR. The third strategy used all S1 conditions ANDed with keywords derived from the medical condition and intervention provided by CT.gov. The manual review was facilitated by a spreadsheet file that included all trial parameters from CT.gov, count of articles retrieved by all three search strategies and a pre-populated link retrieving publications for all search strategies. Since the goal was to determine negative predictive value of the trial-publication link, the manual review was stopped when any relevant not-linked publication was found. Since we started with a list of trials with no linked

publication, existence of such publication classified the trial as FN. If no relevant publication was found, the trial was counted as TN. The resulting NPV was calculated on trial level.

Results

In total, 117,420 trials were registered in CT.gov as of December 2011. Of these, 52,472 were identified as "completed", of which 14,260 had completion dates specified by our inclusion criteria (CT.gov dataset). Using an automated R script that looped through all 14,260 trials in the CT.gov dataset, we identified 2,668 trials with NCT-linked linked publication (search done in January, 2012) and 930 trials with PMID-linked publications (search done in March 2012). In 339 trials, we found publications with both NCT-linked and PMID-linked publications.

Precision of links between Clinical Trials.gov and PubMed

The precision data set was randomly sampled from all trials with publication link (2668+930 trials identified above). The data set also mimicked the proportion of NCT-linked trials to PMID-linked trials (74% of trials with publication links have an NCT-link, 26% of trials with publications have a PMID-link). It consisted of 300 trials with NCT-linked articles (*NCT-linked set*) and 105 trials with PMID-linked articles (*PMID-linked set*). We characterize those separately.

NCT-linked set: In the NCT-linked precision set, there were on average 1.21 (SD: 0.81) articles per trial. The vast majority of the trials, 86% (258/300), had exactly one linked article. Eleven percent (34/300) had two linked articles and the remaining three percent had three or more linked articles (range: 3-11). We manually reviewed a total of 364 articles related to the 300 trials. A total of 361 articles were classified as true positive (TP) and fully related to the trial in question (reporting primary or secondary trial outcomes). We found three TP* articles where the article described a topic related to the trial or used the trial data, but did not directly report on trial outcomes (discussed below). We found no false positive (FP) articles (articles not related to the trial in question). The resulting calculated NCT-link precision was 100% (364 TP or TP* studies out of the total of 364 reviewed articles). Given that 364 trials were evaluated, the probability that the true precision was only 99% was less than 0.026.

PMID-linked set: For the 105 articles with PMID-links, there was an average of 2.3 articles per trial (SD: 3.8) Similar to the NCT-link, the majority of trials (66%; 69/105) had exactly one linked article, 15% of trials had two linked articles and the remaining 19% had 3 or more articles (range 3-27). We reviewed a total of 246 PMID-linked articles. There were 94 false positives articles that were not about the trial in question (typically, they seemed to be background reading for the study). A total of 144 articles were true positive (TP) and reported on the result of the trial. A total of 10 articles were TP* where the article described a topic related to the trial or used the trial data but did not directly report on trial outcomes as defined by the trial registration data. The resulting calculated precision of PMID-links was 62.6% (154 out of 246). A common problem found during the review of the PMID-linked publication for some trials was categorization of references into incorrect reference types. CT.gov provides the PI the ability to specify two types of references: <result_reference> for reporting the trial results and <reference> defined as publications related to the background of the trial.[6] In 12% of trials in our PMID-linked subset (13 out of 105 trials), we found that all the trial's references were false positive: they were not reporting the results of the trial in question and quite often were published prior the trial start date. An example of such trial is NCT00266318, a study of Interferon Alfacon-1 and Ribavirin in Hepatitis C patients. Manual review indicated that those references were most likely mis-categorized as <result_references> instead of background <references>. Moreover, the average count of references in trials where this mis-categorization occurred was relatively high and hence had significant impact on the overall precision. In a sub-analysis, we computed an additional PMID-link precision after we excluded such trials from the PMID-link precision analysis (removing 13 trials where *all* trial references were false positive). This additionally calculated PMID-link precision was 95.7% (154 out of 161) and the remaining count of false positive articles in our sample was seven.

Table 1. Overview of Precision results

Measure	Measure result	Trial count	Article count	TP (TP*) count	FP count
NCT-link precision	100% (364/364)	300	364	361 (3)	0
PMID-link precision	62.6% (154/246)	105	246	144 (10)	92
PMID-link precision (after exclusions)	95.7 (154/161)	92	161	144 (10)	7

Negative Predictive Value

The NPV evaluation dataset included a random sample of 50 studies with no linked publications (either NCT-linked or PMID-linked) from the CT.gov dataset. An automated R script retrieved the necessary registry data and generated a spreadsheet file with article counts and search links for all three search strategies defined above as preparation for manual review of each trial. To illustrate the number of articles returned by the three search strategies we list in Table 2 the total count of articles (all 50 trials combined) and the average number of articles per trial for all three strategies.

Table 2. Overview statistics of the search strategies.

Search Strategy	Total Articles	Mean (SD)	Median	Range
S1: PI name + date range	1764	35.0 (62.0)	20	0 – 419
S2: PI name + date range + PI keywords	365	7.1 (9.6)	2	0 – 38
S3: PI name + date range + CT.gov keywords	420	8.4 (12.7)	3	0 – 54

Compared with the precision evaluation, the manual review of articles for NPV determination dealt with a much higher number of articles, had a greater span of article topics, and required a more time-intensive comparison of trial description to article abstracts or article full-texts. The order of the reviewed search result sets was guided by the article counts of the three search strategies. The review started with the search strategy with smallest count (usually S2) and if no relevant articles were found, we considered the next strategy, concluding with the least restrictive S1 search involving all publications by the trial PI and a trial specific date range criterion. Because the review stopped at any stage once a relevant article was identified, not all 1764 articles returned by the S1 strategy had to be reviewed.

In 28 trials in our NPV dataset, we did not identify any relevant trial result publications (TN). For the remaining 22 trials (44%), at least one unlinked publication was found during manual review (FN). The resulting calculated Negative Predictive Value was 56% (28/(28+22)).

Discussion

Findings

Precision: The NCT-link precision was extremely high and we did not find any false positive publication links. Assigning a wrong NCT to a publication can occur when authors submit the article or when the information is submitted from the publisher to NLM. We think that the high precision is greatly facilitated by the journal review process. As part of submission process for clinical trial type of articles, numerous journals require the authors to provide the reference to trial registration as a requirement for submission. This entry is subject to review by the journal editorial staff as well as reviewers. Submission of this information by the publisher to MEDLINE is also automated so we did not expect to find typographical errors.

The PMID-link precision was not optimal and confounded greatly by what we attribute to mis-categorization of trial references (i.e., incorrectly characterizing trial background references as result references). It was out of scope of our study to investigate if result references are being incorrectly characterizing as background references. Unlike the NCT-link, the PMID-link is not subject to review by journal editors or peer-reviewers. The precision of the PMID-link, however, could be improved by automated methods and even applied retrospectively to existing links using the following rule: a trial result reference publication date must occur after trial start date. In addition, CT.gov recently increased quality assurance (QA) on new trial registrations[1] and result reference basic curation could be added to such QA procedures.

Negative Predictive Value: The NPV result of 56% shows that there are potentially numerous publications that should have been formally linked to their corresponding trials. This is further evidence that existing trial registry and publisher policies may not be fully enforced. Said a different way, our results show that in 44% of the trials with no linked publications (NCT or PMID linked), we were able to find at least one relevant publication via keyword MEDLINE search. This result is comparable to one prior study by Ross et al.[5] that also involved manual MEDLINE search for relevant publications when no linked publication were present. Additionally, they attempted to contact the PI (with low response rate of 37.6%) if manual search did not identify any relevant publications. Their findings were that on top of 14.2% of trials (96/ 677) that had a linked publication, they were able to find a relevant publication in additional 31.8% of trials (215/677).

Implications

The key findings of our study are quantitative measures of the accuracy of links between CT.gov and PubMed. Due to the volume of clinical trials registered and being published, automated ways of linking are highly preferred and to our knowledge, our study is the first to explicitly measure the quality of this link. The linked trial and publication data enable exploring questions such as: How long does it take to publish trial results? How often are trial results published? Do the published reports follow journal requirements such as those of International Committee of Medical Journal Editors (ICMJE)?[9] Is there a publication bias with positive results?

The key utility of trial registration during its initial phase is facilitating recruitment and helping in research portfolio management: e.g., making sure that not too many identical trials are initiated. After the trial completion, the key utility of the trial registry record is in aiding the research enterprise and acting as a pointer to research results.

For linking trials to publications, consideration of the question of boundaries and what makes an article relevant to a trial is also important. Currently such boundaries are not clearly defined by ICMJE. There is a range of possible types of articles that might be linked to trials. There are clear result publications that report on a single study and report on some or all trial primary outcomes. Other relevant publications of this type are interim result articles (e.g., reports of five-year interim results in a planned ten-year study), articles reporting on two similar studies at the same time (e.g., organized by the same industry sponsor), or follow-up studies on original trial participants or their samples.

There are also publications that may only marginally report on the trial results but may represent the only available publication with at least some insights generated by the study. This may happen in studies where a full article may be difficult to publish. Examples are editorials that briefly mention the trial in question and include very limited trial results. We used our special category of true positive (TP*) to explore possible boundaries or policy recommendations for article authors linking an NCT identifier to their publication during submission or revisions (NCT link) or enumerating result publication within a trial register (PMID link). An example of a TP* is an article entitled "A comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data" (PMID:19472307), which was linked to a trial entitled "Antidepressant Therapy for Bipolar II Major Depression" (NCT00641927). The linked statistical article describes a novel statistical method and uses the trial dataset; however, it does not report at all on the primary or secondary clinical outcome of the trial (depression treatment is not mentioned in the article abstract and is only evident from the full text). Nevertheless, the authors we correct to link the publication to the trial.

Another important consideration is the number of linked articles and the high significance of at least one linked article. Researchers seeking individual trial results can subsequently use a citation network of that single linked publication to retrieve other relevant unlinked publication. Examples of search engines offering "cited-by search" are Google Scholar, Thomson Reuters's Web of Knowledge or CrosRef's Cited-by service. The single relevant publication can be used to seed further searches and can reduce greatly the search space compared to just trial keywords and date range limit. However, when considering link cardinality (one trial to several articles or one article to several trials), it is important re-state the finding of 1:1 cardinality in 86% of the trials in the NCT-linked sample (The link cardinality figure in the PMID-linked sample is biased by the mis-categorization phenomenon). A deeper exploration of the trial-article link cardinality was out of the scope of this study and we plan to investigate it in a follow-up analysis covering a much larger sample of trials from CT.gov.

Finally, based on the results above, it is possible to recommend for future manuscript authors the type of trial-publication link (NCT vs. PMID) that would be preferred for future analyses of trials' metadata. In terms of current usage frequency, the NCT-link is almost three times more common than the PMID-link, and an NCT-link is required by policies of numerous journals for clinical trial articles. Our results show that the NCT-link is also more precise (most likely due to the editorial and peer review processes) and it would be our recommended way of linking trials with publications.

Limitations

Our study has several limitations. First, we analyzed only one trial registry, ClinicaTtrials.gov. Several other registries (e.g., the ISRCTN registry) are covered by the MEDLINE secondary identifier field; however, CT.gov is the largest trial registry by volume and provides the most comprehensive access to its data. Second, we only looked at publications indexed in PubMed. We think that researchers potentially searching for relevant publication would

also most likely use PubMed exclusively to search for high quality publications. Third, in determining precision, we analyzed NCT-links and PMID-links separately, and it was out of our current study scope to investigate how they overlap (for example, how often the PMID-link provides additional articles not covered by the NCT-link). We also did not pursue assessment of inter-rater agreement and each trial was reviewed by only one reviewer. Finally, our NPV result is, to a degree, dependant on the completeness of the three search strategies used to identify relevant publications. However, the same keyword-based search was used by prior studies.[5] The key restricting component is the principal investigator's name, otherwise the number of returned articles makes manual review of search results clearly not feasible. We believe that the exclusion of the PI from the authorship of all relevant publications is probably uncommon.

Conclusion

With the increased sophistication of the MEDLINE database, advanced querying using newly created MEDLINE attributes is becoming more common. Links between a clinical trial registry and PubMed can facilitate advanced insights into the world literature and the research enterprise. We investigated the accuracy of such links and provide some recommendations for improving them. We think that an accurate and complete "*trialome*", as a collection of well organized data about human clinical trials, represents an important asset for several domains: (1) translational science to guide research prioritization, (2) public health to support creation of accurate recommendations and clinical guidelines, and (3) individual patient care to support evidence-based medicine. We have also identified several informatics challenges of current data and data access relevant to clinical research informatics efforts to better support the overall research enterprise.

Acknowledgments and Disclaimers

This work has been supported by intramural research funds from the NIH Clinical Center and the National Library of Medicine. Mention of particular commercial products does not imply endorsement of those products.

Glossary

NCT-Link is one type of trial-publication link that is specified at manuscript submission time, in which a publication describing a clinical trial provides the trial ID or "NCT number" assigned by the trial registry and the trial ID is clearly presented in the article abstract. This information is extracted during PubMed indexing and becomes an article attribute within MEDLINE database.

PMID-Link is a second type of trial-publication link that is specified by the trial PI when he or she updates the trial registry data. It is an optional trial attribute within the ClinicalTrials.gov database, stored as a parameter called <result reference/PMID> that contains the PubMed article ID that describes the results of the trial.

TP is a subcategory of true positive articles used during our precision evaluation to categorize the article as centrally related to the trial in question. Such an article directly reports on trial outcomes as defined by the trial registration data (trial description and trial primary and secondary outcomes).

*TP** is a subcategory of true positive articles used during our precision evaluation to categorize the article as marginally related to the trial in question. Such an article typically describes a topic related to the trial (or used the trial data) but does not directly report on trial outcomes as defined by the trial registration data (trial description and trial primary and secondary outcomes).

Trial with linked publication is a trial that has at least one linked MEDLINE article via either NCT-link or PMID-link.

Trial with no publications is a trial which has no linked publications and no publications can be found by searching MEDLINE using any search strategy or by contacting the trial principal investigator.

Trial with only not-linked publications is a trial for which related results publications exist, but none of them are linked to the trial via the NCT-link or the PMID-link.

References

1. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database--update and key issues. N Engl J Med. 2011 Mar 3;364(9):852-60.

2. FDA Modernization Act, 1997, section 113.
3. Marusic A, Huic M. Registration of clinical trials still moving ahead--September 2008 update to Uniform Requirements for Manuscripts Submitted to Biomedical Journals. Croat Med J. 2008 Oct;49(5):582-5.
4. Ramsey S, Scoggins J. Commentary: Practicing on the Tip of an Information Iceberg? Evidence of Underpublication of Registered Clinical Trials in Oncology (vol 13, pg 925, 2008). Oncologist. 2008 2008;13(10):1128-.
5. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.Gov: a cross-sectional analysis. PLoS Med. 2009 Sep;6(9):e1000144.
6. ClinicalTrials.gov Protocol Data Element Definitions. Available from: http://prsinfo.clinicaltrials.gov/definitions.html.
7. Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. BMJ. 2012;344:d7373.
8. MEDLINE/PubMed XML Element Descriptions and their Attributes. Available from: http://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html#databanklist.
9. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: . Available from: http://www.icmje.org/urm_main.html.