

Trial Bank Publishing: Phase I Results

Ida Sim, Simona Carini, Ben Olasov, Sophia Jeng

Program in Biological and Medical Informatics, University of California San Francisco, CA, USA

Abstract

Randomized clinical trials (RCTs) are an important source of evidence for clinical practice, but finding and applying RCT reports to care is time consuming. Publishing RCTs directly into machine-understandable "trial banks" may allow computers to deliver RCT evidence more selectively and effectively to clinicians. Methods: Authors of eligible RCTs published in JAMA or the Annals of Internal Medicine between January 2002 and July 2003 were invited to co-publish their trial in RCT Bank, an electronic knowledge base containing details of trial design, execution, and summary results. Trial bank staff used Bank-a-Trial, a web-based trial-bank entry tool, to enter information from the manuscript into RCT Bank, obtaining additional information as necessary from the authors. Results: The author participation rate rose from 38% to 76% after the first co-published trial was available as an example. Seven diverse RCTs are now co-published, with 14 in progress. Conclusions: We have demonstrated proof of concept for co-publishing RCTs with leading journals into a structured knowledge base. Phase II of trial bank publishing will introduce direct author submission to RCT Bank.

Keywords:

Publishing, evidence-based medicine, randomized trials, databases, knowledge bases, meta-analysis

Introduction

Randomized clinical trials (RCTs) are one of the least-biased sources of evidence for the practice of medicine, but substantial gaps often exist between every day practice and "best practice" as defined by research evidence. For example, evidence from the 1980s showed that beta-blockers reduce mortality in myocardial infarction (MI) survivors [1], yet even in the mid 1990s, only 21 to 77 percent of eligible heart-attack patients were receiving post-MI beta-blockers [2, 3].

With over 10,000 RCT reports indexed annually in Medline, computers are desperately needed to help physicians translate RCT evidence into practice more efficiently. However, RCT findings are published only as text articles that are of limited machine understandability, such that computers are effectively illiterate of the RCTs they are supposed to help clinicians apply to care. Rather, RCT findings should be available in machine-understandable "trial banks" that decision support systems can use to drive recommendations for evidence-based practice. Trial banks should include newly published RCT findings to ensure that decision support systems can be evidence adaptive [4]. Thus, we propose a new model of clinical scientific publishing

called trial bank publishing, in which RCTs are reported as both prose articles and as entries into a trial bank [5].

The Trial Bank Project captures RCTs that are published by *JAMA* or the *Annals of Internal Medicine* (*Annals*) into a trial bank called RCT Bank. Given the high stakes involved in manuscript review and publication at top-tier journals such as *JAMA* and *Annals*, and given the conservatism of medical publishing, we chose a phased approach to introducing trial bank publishing. In Phase I, reported in this paper, trial bank staff enter trial information into RCT Bank, using information from the manuscript and from queries to the authors. In Phase II, authors themselves will use Bank-a-Trial to enter their trials into RCT Bank at the time they submit their manuscripts to the journal.

Like the publishing of genomic sequences into GenBank [6], trial bank publishing embodies the principle that scientific knowledge should be disseminated in the form that best facilitates its use; simply publishing science in electronic text is not good enough anymore [7].

Methods

Trial Bank System

The Trial Bank system (Figure 1) captures RCT information into RCT Bank, a knowledge base designed to support systematic reviewing and evidence-based practice. Data can be securely entered into RCT Bank using Bank-a-Trial, a web-based data entry tool, and can be browsed over the web using RCT Presenter.

RCT Bank is implemented as an Ocelot [8] frame-based knowledge base in Allegro Common Lisp v6.2 (Franz, Inc.) on SunOS 5.8. Bank-a-Trial and RCT Presenter are generated using the Common Lisp Hypermedia Server (CL-HTTP) [9]. RCT Bank codes clinical concepts in the Unified Medical Language System (UMLS) [10]. Password-based user authentication is handled via CL-HTTP functions. RCT Bank can be accessed programmatically by outside users via a Java- or Perl-based API.

RCT Schema

RCT Bank, as modeled in its data schema RCT Schema, should contain all the trial information that humans or computers will need to retrieve, summarize and apply RCT evidence to clinical care. The canonical method for summarizing and applying RCT evidence to clinical care is *systematic reviewing* [11]. Systematic reviewing involves retrieving a set of trials that explore the same clinical question, examining the quality of the individual trials, investigating any clinical and statistical differences across the trials, and then, when appropriate, statistically combining the

quantitative results to increase the statistical power for detecting intervention effects [12].

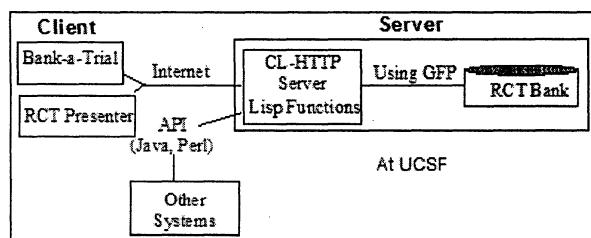


Figure 1 - The Trial Bank System Architecture

Thus, to support evidence-based practice, trial bank entries should contain all the trial information necessary for systematic reviewing. That is, systematic reviewing should be the *target task* for trial banks, and the information requirements of systematic reviewing should drive the data modeling for trial banks. In prior work, we have defined a task analysis of systematic reviewing that specifies what trial banks ought to contain (see <http://rctbank.ucsf.edu/tasks/tasks.html>).

Based on this task analysis of systematic reviewing, RCT Schema captures details about the protocol, execution, follow-up, and summary results of the most common types of RCTs at any stage of implementation, from planned to fully completed. It does so with a class hierarchy that is 7 levels deep, with 189 frames and 616 unique slots. There are an average 10.1 slots per frame, and 188 of the 616 slots (30.5%) take other frames as values. Thirteen frames (7%) have multiple parents.

RCT Schema does not yet capture factorial, crossover, or follow-up trials, or trials with cluster randomization.

Bank-a-Trial

It consists of over 135 dynamically generated web pages that allow users such as clinician investigators or systematic reviewers to enter trials into RCT Bank (Figure 2). Users are assumed to be familiar with clinical trial terms (e.g., primary hypothesis, intention-to-treat analysis), but are not expected to have any special computer science or knowledge modeling expertise.

After securing authorization to edit an existing trial or add a new one, users can enter information about that trial's administration, background, design, entrance criteria, randomization, interventions, outcomes, or conclusions in any order. Once they have described the study interventions and outcomes, users can then enter details about subject enrollment, follow-up, and compliance. Finally, after subject enrollment and follow-up have been described, the study results can be entered. The order of data entry is constrained in the way just described, to allow the customization of subsequent pages. For example, the pages for collecting study results are customized to previously entered descriptions of the study's interventions, outcomes, and subject enrollment (including subgroup definition and size).

Most data fields are either free text or pick lists. Data integrity checks are being added, but users are allowed to enter inconsistent data (e.g., that some subjects are not accounted for at follow-up), if that is the most accurate data available. While we work with authors to resolve such discrepancies, those that remain un-

resolved at the time of RCT Bank publication are described in the entry notes for that trial.

For each entrance criterion, intervention, outcome, and subgroup population, users are asked to select a single descriptive UMLS term. Bank-a-Trial prompts users to submit one or more (sub)strings, returns all matching UMLS terms, and, when available, provides a definition of the selected term to help the users ensure that they have selected the most appropriate term. Bank-a-Trial then stores the original search string(s), the String Unique Identifier (SUI) of the UMLS term string that the user selected, as well as the Concept Unique Identifier (CUI) of the corresponding UMLS preferred term. UMLS terms are stored in a MySQL® database accessed using Allegro ODBC.

RCT Presenter and SecurePresenter

RCT Presenter and SecurePresenter allow browsing of public and access-controlled RCT Bank entries respectively. Presenter pages are generated dynamically so that updates to trial-bank entries are reflected directly on them. RCT Presenter lets the users retrieve from RCT Bank the trial(s) of interest based on criteria such as intervention, outcomes, length of follow-up, and population. Users can request specific trial information, linear short or long summaries of a trial, or they may browse the information using hyperlinks and other navigational aids (Figure 3). Users can also view trial reporting and quality assessment guidelines (e.g., CONSORT) that are hyperlinked directly to the relevant information for the selected trial. The generation of arbitrary tables of information across selected RCT Bank entries is currently under development.

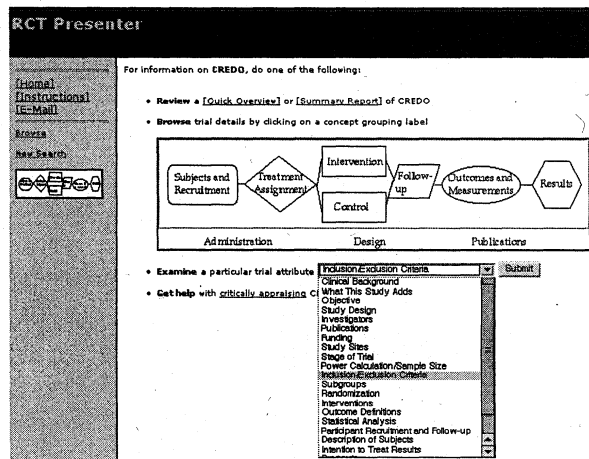


Figure 3 - RCT Presenter for Browsing a Trial

Phase I Trial Bank Publishing

Our trial bank publishing collaborators are *JAMA* and *Annals*, two top-tier general medical journals. Our collaboration with *JAMA*, which started in January 2002, is better established and has generated the vast majority of the trials currently in RCT Bank. While other journals have expressed an interest in co-publishing their articles in RCT Bank, we have focused our efforts only on establishing a working collaboration with *JAMA* and *Annals*. Our procedure for Phase I trial bank publishing is similar for *JAMA* and *Annals*.

Choose Section to Edit

Section Brief	Description	Action
Administration	Trial title, stage, dates, and investigators, study sites, funding, ethics, and description of trial committees	Edit
Background	Text description of study background, objectives, and rationale	Edit
Design	Design of trial, statistics used, details of randomization and allocation concealment, and subgroup information	Edit
Entrance Criteria	Inclusion and exclusion criteria	Edit
Interventions	Description of interventions and cointerventions, and details of treatment masking and administration	Edit
Outcome Variables	Definitions of outcome variables, baseline characteristics, side effect variables, and details of outcome analysis, assessment, and measurement	Edit
Enrollment	Recruitment, screening, and enrollment of subjects, and number of subjects in each intervention group	Need to complete: Experimental Arms Comparison Arm
Followup and Compliance	Followup of subjects, compliance, and crossovers	Need to complete: Experimental Arms Comparison Arm Recruitment Flowsheet
Results	Quantitative study results	Need to complete: Experimental Arms Comparison Arm Recruitment Flowsheet
Conclusions and Publications	Discussion of study's limitations and conclusions, and citations to trial publications	Edit

[Click here to select another trial for editing](#)

Figure 2 - Bank-a-Trial Main Menu

For simplicity, we describe only the *JAMA* procedure. Authors submit RCT manuscripts and are peer-reviewed as usual. Trial Bank staff receive a pre-publication copy of the journal and determine whether published RCTs can be captured by RCT Schema (i.e., they are not factorial, crossover, follow-up, or cluster randomized trials). First authors of eligible RCTs then receive an e-mail invitation from a *JAMA* deputy editor (Dr. Drummond Rennie) and the Trial Bank Project principal investigator (Dr. Ida Sim) to allow co-publication of their trial in RCT Bank. The e-mail includes a link to the RCT Presenter web site, so that authors can browse previously co-published trials. For participating RCTs, Trial Bank Project staff enter data directly from the *JAMA* manuscript into RCT Bank, requesting additional information from the authors as necessary in cases of missing or discrepant information. When a trial's RCT Bank entry is completed, the authors are notified and a hyperlink to the trial's record in RCT Presenter is added at the end of the article's full-text version at *JAMA*'s website.

Trial bank publishing is guided by a Trial Bank Advisory Board consisting of six internationally renowned editors, epidemiologists, and clinical researchers.

Results

Author Participation

From January 16, 2002 to July 31, 2003, 108 RCTs were published in *JAMA*. Of these, 54 were excluded due to modeling limitations of RCT Schema (Table 1). Only one *Annals* author, a convenience sample, was invited.

The overall participation rate was 64%. Initially, the rate was 38% (8 out of 21 invited), including the first five invitations that

were extended weeks after the RCT had been published, and which did not garner any response from the authors. After the first co-published trial was available as an example for invited authors to browse, the participation rate rose to 76%.

Table 1: Flowchart of Participating Trials

Reason	JAMA	Annals	Total
RCTs published	108		108
Excluded	54		54
Invited	54	1	55
Accepted	34	1	35 (64%)
Invited after RCT Bank example	34		34
Accepted after RCT Bank example	26		26 (76%)

Trials Co-Published

RCT Bank currently has 7 co-published trials (6 from *JAMA* and 1 from *Annals*) publicly available via RCT Presenter. These trials are diverse in their clinical domains, and intervention, outcome, and result types (Table 2).

Other trial attributes that we were able to capture in RCT Bank include participant dropout at any stage of recruitment and follow-up (e.g., after randomization but before any intervention was started), early stoppage, protocol changes during trial execution, and how accurate subjects were at guessing which intervention arm they had been assigned to. All of this information is relevant to any rigorous interpretation or application of trial results, as detailed in our task analysis of systematic reviewing.

We contacted all authors for additional information. For most authors, the inquiries consisted of 2 to 3 requests for clarifications (e.g., for which outcome was the study powered), although

1 study required 8 detailed questions to resolve. Three studies had data discrepancies: subject numbers or result values did not match in the tables or figures and the text. Authors responded in 1 to 4 weeks, often asking their statisticians to assist. Two authors did not provide complete responses to information requests on non-critical items. No author withdrew their participation because of data requests or any other reasons.

Table 2: Characteristics of RCT Bank Trials

Characteristic	Examples
Clinical domains	Cardiology, Radiology, Geriatrics
Intervention Types	Procedures (thrombolysis), Single and Multi-step Drugs (aspirin, warfarin), Counseling, Multiple interventions in one arm (8 kinds of fall prevention interventions)
Outcome Types	Dichotomous, continuous, univariate, multivariate, survival, regression, scored instruments (e.g., Wechsler Memory Scale)
Result Types	Intention-to-treat, efficacy analysis, subgroup analyses

Trial bank staff required from 8 to 20 hours to enter a trial completely, with most trials taking 10 to 15 hours. On average, only 10 to 20% of the time was needed for actual data entry; the rest of the time was needed for reading the manuscript — often between the lines — for information. For example, one trial was powered for an outcome that was not reported in the study, and the reporting of outcome results was sometimes inconsistent with the description of those outcomes.

Trials In Progress

14 trials are in progress: undergoing data entry, awaiting additional information from authors, or awaiting modeling changes to RCT Schema to accommodate trial features. For example, several of the multi-armed trials report pairwise comparisons between arms, a result type that requires additional modeling in RCT Schema, and corresponding changes to Bank-a-Trial and RCT Presenter.

For one of the trials currently in progress, we invited the authors to submit their trial to RCT Bank using Bank-a-Trial. The trial's statistician did so, and found Bank-a-Trial very easy to use. As a first time user of Bank-a-Trial, she was able to enter the bulk of the information regarding the trial without any special training or assistance in under 3 hours.

Website Usage

RCT Presenter, the website for browsing co-published trials in RCT Bank, can be accessed without charge either directly at <http://rctbank.ucsf.edu/Presenter/>, or from a link at the end of the trial's full-text article on the *JAMA* or *Annals* website, which is available only to journal's subscribers. RCT Bank trials that are also in ClinicalTrials.gov will have reciprocal links with RCT Presenter; however, there are no shared trials at present. Other known links to RCT Presenter include OpenClinical [13].

Traffic to RCT Presenter from January through August 2003 is shown in Figure 4. An average of 180 unique IP addresses per month visited RCT Presenter, excluding visits from web crawl-

ers and other automated visitors. Of North American visitors, 56% of IP addresses were from .com and .net domains, 20% from .edu, 3% from .org and 21% from other domains.

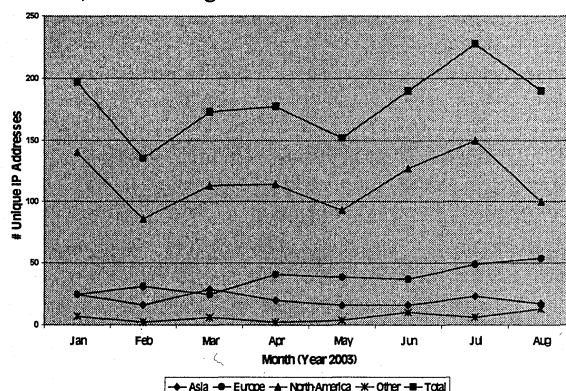


Figure 3 - Figure 4 – RCT Presenter Visitor Traffic

Other Usages

In addition to using the Trial Bank system for trial bank publishing, we have four external users from academia and industry building their own domain-specific trial banks. Several research groups are accessing the RCT Bank API to programmatically retrieve detailed trial information for use in their decision support.

Discussion

Phase I of the trial bank publishing project has successfully shown proof of concept for co-publishing RCTs with leading medical journals into a structured knowledge base. Both *JAMA* and *Annals* have been accommodating: they have modified editorial procedures to allow us to contact authors and to review manuscripts prior to publication. Authors, while initially unresponsive to our invitations, were highly responsive — sometimes very enthusiastically — once an example of trial bank publishing was available for them to view. Data entry using Bank-a-Trial required only a modest expenditure of time, especially when compared to the amount of time required to prepare a manuscript for publication. Based on very early experience, it appears that untrained users have little difficulty using Bank-a-Trial to enter trial bank information. Traffic to RCT Presenter is modest at present, as the website is still new and has not yet been actively promoted.

RCT Bank serves as an open access repository of RCT evidence for decision support systems to use as they deliver recommendations for evidence-based practice. RCT Bank differs from existing trial databases such as ClinicalTrials.gov [14], the Cochrane Controlled Trials Register (CCTR) [15], the PDQ database [16], and other trial registries [17] in a crucial way: RCT Bank can capture all the 147 unique information items necessary for systematic reviewing, the canonical method for trial interpretation and application, while other trial databases contain only sufficient information for recruiting subjects, or for determining a trial's existence. Thus, trial bank publishing into RCT Bank could significantly enhance the informatics infrastructure for evidence-based decision support systems.

We encountered both computing and sociotechnical challenges in this first phase of trial bank publishing. There was a surprising amount of heterogeneity in trial design requiring modifications to RCT Schema (e.g., subgroup outcomes assessed at different times from main intervention groups, a trial powered on a secondary outcome, dropouts after randomization that were not included in either the intention-to-treat or efficacy analyses). With these modifications, we have been able in recent months to capture all features of 2-armed trials. Additional modeling is still required to capture some aspects of 3-armed trials (e.g., pairwise comparisons) for much of the backlog of trials in progress.

The predominant sociotechnical challenge we faced was to devise a way of integrating trial bank publishing with each journal's manuscript review, acceptance, and publication procedures, while maintaining the highest levels of confidentiality, not overburdening either the editors or the authors, and protecting intellectual property rights. Both journals are supportive of RCT Presenter being open access: JAMA initially placed the link to RCT Presenter on their articles' abstract pages, but the abstract page is free to the public and would have pointed all viewers to the RCT Presenter site where all the published trial information (and more) is available. Subsequently, they moved the link to the subscriber-only full-text page. Although neither journal participates in the open access PubMed Central repository, they do not see open availability of their trial information in RCT Presenter as an intellectual property threat. Copyright to the article is maintained by the respective journal, while the University of California holds copyright to the RCT Presenter version of each trial.

A formal evaluation of RCT Presenter is under way to assess clinician and systematic reviewer satisfaction with trial bank publishing versus traditional journal publishing. If the evaluation is favorable, we will be better able to convince journals and authors to proceed to Phase II of trial bank publishing, in which authors themselves will submit their trials to RCT Bank in conjunction with submitting their manuscript. Peer review will then be performed on both the manuscript and the trial bank entry using SecurePresenter. Accepted trials would be released on RCT Presenter; rejected trials could be submitted to another journal or be expunged.

As trial bank publishing takes hold, RCT Bank should become an increasingly valuable resource of shared, machine-understandable, recently published RCT evidence that decision support systems can use to support evidence-based practice. Ongoing challenges include sustainable approaches to accruing entries into RCT Bank, accrual of RCTs from other journals and sources (e.g., as part of the National Institutes for Health data sharing initiative), and demonstration of effective RCT Bank use by other information and decision support systems.

Acknowledgments

Supported by LM-06780, and Dr. Sim's United States Presidential Early Career Award for Scientists and Engineers.

References

- [1] Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of

the randomized trials. *Prog Cardiovasc Dis* 1985;27(5):335-71.

- [2] Soumerai SB, McLaughlin TJ, Spiegelman D, Hertzmark E, Thibault G, Goldman L. Adverse outcomes of underuse of beta-blockers in elderly survivors of acute myocardial infarction [see comments]. *JAMA* 1997;277(2):115-21.
- [3] Krumholz HM, Radford MJ, Wang Y, Chen J, Heiat A, Marciniak TA. National use and effectiveness of beta-blockers for the treatment of elderly patients after acute myocardial infarction: National Cooperative Cardiovascular Project. *JAMA* 1998;280(7):623-9.
- [4] Sim I, Gorman P, Greenes R, Haynes R, Kaplan B, Lehmann H, Tang P. Clinical Decision Support Systems for the Practice of Evidence-Based Medicine. *J Am Med Inform Assoc* 2001;8(6):527-534.
- [5] Sim I, Owens DK, Lavori PW, Rennels GD. Electronic Trial Banks: A Complementary Method for Reporting Randomized Trials. *Med Decis Making* 2000;20(4):440-450.
- [6] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res* 2003;31(1):23-7.
- [7] Sim I, Sanders GD, McDonald KM. Evidence-based practice for mere mortals: the role of informatics and health services research. *J Gen Intern Med* 2002;17(4):302-8.
- [8] Karp P, Chaudhri V, Paley S. A Collaborative Environment for Authoring Large Knowledge Bases. *Journal of Intelligent Information Systems* 1998:In press.
- [9] Common Lisp Hypermedia Server (CL-HTTP), version 70.23. 2003. Cambridge, MA.
- [10] Unified Medical Language System Fact Sheet. National Library of Medicine. Available at: <http://www.nlm.nih.gov/pubs/factsheets/umls.html>. Accessed September 10, 2003.
- [11] Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126(5):376-80.
- [12] Chalmers I, Altman D, editors. Systematic Reviews. London, UK: BMJ Publishing Group; 1995.
- [13] OpenClinical: Knowledge Management for Medical Care. OpenClinical. Available at: <http://www.openclinical.org/>. Accessed September 10, 2003.
- [14] ClinicalTrials.gov. National Institutes of Health. Available at: www.clinicaltrials.gov. Accessed Sept. 16, 2002.
- [15] Collaboration C. The Cochrane Library [CDROM]: Update Software; 2002.
- [16] Cancer.gov. National Cancer Institute. Available at: http://www.nci.nih.gov/search/clinical_trials/. Accessed November 18, 2002.
- [17] Easterbrook PJ. Directory of registries of clinical trials. *Stat Med* 1992;11(3):345-59.

Correspondence

3333 California St. Suite 435Q, San Francisco, CA. USA
94143-1211
sim@medicine.ucsf.edu