Study-based registers of randomized controlled trials: Starting a systematic review with data extraction or meta-analysis

Farhad Shokraneh1*, Clive Elliott Adams1

1Cochrane Schizophrenia Group, the Institute of Mental Health, a partnership between the University of Nottingham and Nottinghamshire Healthcare NHS Trust, Nottingham, United Kingdom

Abstract

Introduction: Despite years of use of study-based registers for storing reports of randomized controlled trials (RCTs), the methodology used in developing such registers/databases has not been documented. Such registers are integral to the process of scientific reviewing. We document and discuss methodological aspects of the development and use of study-based registers. Although the content is focused on the study-based register of randomized/controlled clinical trials, this work applies to developers of databases of all sorts of studies related to the human, animals, cells, genes, and molecules.

Methods: We describe necessity, rationale, and steps for the development, utilization and maintenance of study-based registers as well as the challenges and gains for the organizations supporting systematic reviews of the published and unpublished literature.

Conclusion: The ultimate goal of having a study-based register is to facilitate efficient production of systematic reviews providing rapid, yet accurate, evidence for the decision-makers. We argue that moving towards study-based registers is an inevitable welcome direction and that infrastructures are ready for such movement.

Introduction

Emergence of specialized organizations creates needs for the specific information for decision-making. Developing and maintaining specialized registers or databases is a valuable tool for achieving organizational goals by providing information to support such decision-making.1,2 Specialized bibliographic registers save time by allowing specific and sensitive searches for records of documents of high relevance to the research question.3

The Cochrane Collaboration (1993 – present) is an international non-for-profit network of, amongst others, healthcare researchers, practitioners, and patients. This Collaboration works to summarize and synthesize available evidence to support informed decision-making in the healthcare sectors. The Collaboration’s main activity is producing high-quality systematic reviews of randomized controlled trials (RCTs).4 However, over the last two decades after its foundation, there are still only 6230 reviews published on the Cochrane Library (June 2017), and most are in need of update. Despite enormous efforts of thousands of researchers, there remain hundreds of thousands of clinically valuable randomized trials not summarized within reviews. Best evidence to support the health care decision-making is being wasted. It is not an impossible task to provide good coverage of all clinically useful and best evidence of the effects of health care in any one sub-specialty but it is a large and difficult task. Since a systematic review requires a long process including literature searching, de-duplication of search results, screening the search results, obtaining the full text of the reports, putting the publications from the same study together (studiﬁcation), data extraction and meta-analysis, it has become clear that study-based registers of randomized trials are integral to this task shortening...
the process for the reviewers to start a review with data extraction or meta-analysis.\textsuperscript{3,6}

From the very start the Cochrane Collaboration created a specialized register of relevant literature – now called the Cochrane Central Register of Controlled Trials (CENTRAL) – and has disseminated this in Cochrane Library.\textsuperscript{7,9} CENTRAL is now the largest bibliographic reference-based database of reports of randomized trials.\textsuperscript{10,11} It is an amalgamation of the individual Cochrane groups’ registers which, in turn, are developed from biomedical bibliographic databases such as EMBASE, MEDLINE and other sources such as conference proceedings.\textsuperscript{12} In such databases, the publications and the reports from the same study are not linked together and putting the salami of the study back together required massive efforts.

Slow production of update systematic reviews because of long process (6230 reviews in 20 years) and salami publication of trials were two main problems leading in the development of study-based registers. In recognition of the value of study-based registers for more efficient review production the methodology for development, utilization and maintenance of such register have been documented in this methodology paper.

Objectives

- To search the literature for relevant papers on study-based registers.
- To report the rationale methods of development and challenges of study-based registers for which relevant documentation seems remarkably sparse.
- To share more than two decades of practical experience in the creation and maintenance of study-based registers of biomedical literature.

Search methods

We ran a search to find relevant literature and to ensure the novelty of the current paper. Since the terminology in this topic is standard, we used the following search strategy on MEDLINE (1946 to Search Date) and EMBASE (1974 to 2017 Week 34) via Ovid SP and updated this search on August 29, 2017:

\texttt{Search Strategy: ("Study-Based" adj (Register* or Database*)),ti,ab.}

We also searched all conference abstracts presented in Cochrane meetings. We did not identify any full published paper on the study-based registers but did identify many conference abstracts. We then contacted Cochrane Information Specialists and followed web searching to find all possibly relevant documents for this overview.

Specialized registers

1. Reference-based registers

Within a reference-based register (database) each record usually behaves as a separate independent entity. There is one record for each ‘reference’ (also referred as ‘citation’, ‘publication’ or ‘report’). However, a study may have several reports or publications and hence, several records in the register.\textsuperscript{13} For instance, the researchers may conduct one study but report it in several references such as conference abstracts, a dissertation, a poster, journal papers, an online trial registry record in “ClinicalTrials.Gov” and so on. Although the references are different in the format they all present some or all data from a single study and each is listed separately within the reference-based register. A reference-based register is the least resource-intensive and simplest register to assemble.\textsuperscript{13}

A reference record in a reference-based register may consist of bibliographic information, abstract and indexing fields. This record may also contain a link to the full report of the reference or be the full report in itself (i.e. conference abstract).\textsuperscript{14,15} In such a register, the reference record may sometimes contain another link to other references of the same study,\textsuperscript{16} or a study name or unique identifier, such as online trial registry number.\textsuperscript{17} Such reference records are the backbone of all the major bibliographic biomedical databases such as MEDLINE and communication between such databases and reference management packages is usually easy. This type of register may represent the totality of publishing activity, but, because of the ‘one-to-many’ issue of so-called ‘salami’ publication of one study, the register is not an accurate representation of the total research activity.

2. Study-based registers

Conducting one study, the researchers usually published their data in more than one paper.\textsuperscript{18} It means that one study might have several reports. To represent the whole data, we should find, use and cite the whole study rather than one paper or one reference of the study. For those working in systematic reviewing – summarizing research activity rather than publishing activity – there is the necessity to work at the study level instead of the reference level. Every time a systematic review is conducted a study-based register of some sort is created. Recognizing the wastage involved in disassembling these at the end of each review, some Cochrane groups now maintain study-based registers in which all references of one study are linked to a single meta-record called the ‘study record’ (also referred to as the ‘trial record’).

The idea of linking related records to one meta-record is not a new concept in information science\textsuperscript{19} but it is the cornerstone of modern relational databases. At the most basic level, a study-based register consists of study records each of which links to all available references of each study.\textsuperscript{20} Further value may be added to the study record by having it contain structured study-specific meta-data such as data about who were the participants, what are healthcare Problems, what are Interventions and Comparator, and what Outcomes have been measured so-called PICO framework.\textsuperscript{21,22} These metadata make it possible to run a very specific search for the particular
intervention relevant to a certain healthcare condition and retrieve all the related trials – and no more – at the click of a button. All the data and their associated meta-data are stored in data tables (Figure 1). One meta-record may also contain information about all relevant references. Even though a reference may have contained similar PICO meta-data there is no implicit guarantee that data in the one reference record accurately represents all activity within the study. A single reference often reports a subset of activity within the study.

The study record can also contain accurate controlled vocabulary describing the comparisons within the study (e.g. drug X versus drug Y for condition Z) with the potential for direct linking to the end review(s). Finally, the study record may also contain all tabulated data extracted from the full-text reports or the full primary dataset (Table 1).

Rationale

The ‘unit of currency’ of systematic reviews of healthcare is the study – not a single reference or a single published paper. The undertaking of a systematic review necessitates the creation of some sort of study-based list or register whether it be by the reviewers or Information Specialist. If pre-prepared by the Information Specialist the final product can be more accurate, save the time of other researchers, and has the potential to be re-used to avoid duplication of effort in the future. Certainly, if the single reference is considered the main unit for systematic reviews, erroneous double or even triple counting of data from various references of a single study is a real danger (Box 1). The use of study-based registers decreases this danger – although it does not make it impossible.

1. Reference-based register or study-based register?

Some research groups prefer to keep their trials register at a reference level. However, many in systematic reviewing research have made (what we feel to be) the inevitable move to becoming study-based. In reality, one has to underlie the other. The study-based register should sit over a clean repository of references.

Moving to a study-based register depends on the research teams’ ethos and policies, planning and available resources. Holding a reference-based register will save search and de-duplication time. However, a study-based register allows Information Specialists not only to save that searching and de-duplating time but also can benefit authors of reviews by providing study records with multiple references already organized at study level. Study-based registers also have the capacity to save author-time by supplying records in which qualitative and numerical data are extracted and referenced.

2. Comparing functionality

Because of the high relevance of screened references to the methodology and healthcare condition, having a specialized reference-based register does increase the sensitivity and specificity of searching for reviews. However, the processes of searching the register, obtaining full-text reports of references, screening search results, checking each reference based on inclusion criteria, assigning references to relevant studies, concatenating data from the same study, and coding information/data to link references to study take time – mostly reviewers’ time. A study-based register has, to a greater or lesser extent, already undertaken at least some of those tasks moving the expense down the line to Information Specialists who are more skilled in such types of information management.

Making the move

The Cochrane Dementia and Cognitive Impairment Group spent 3 years converting their specialized register...
from reference-based to study-based format.\textsuperscript{37} We have
found no other estimates of duration of the switch for other groups in health care. This may seem
discouraging but the initial shift is a matter of days. Fully utilizing
functionality does, however, take longer. Any initial
investment is offset by the recognition of the daily
waste occurring in the systematic reviewing process in
discarding efforts of reviewers when supplying data to the
next generation of researchers. The past and considerable
efforts in concatenation into study and data extraction are
routinely not re-issued. Repeating such efforts of merging
(sometimes) hundreds of references into one study is not
time-efficient. This merging and extraction of data should,
of course, be entirely transparent to future users (study
records can potentially hold such tracking information).
As already discussed, a relational database such as
study-based register has the capacity to provide one-to-
many relationships so that one study with many reports,
but, in addition, conversely, one report referring to data
from many studies could be linked and managed easily
(many-to-one relationships).

1. Developing a register
The development process of specialized registers is
described in Table 2. For study-based registers, guidelines
should clarify what fields are required.\textsuperscript{38,39} As the study

\begin{table}
\centering
\caption{A study record containing the metadata, data extracted from a trial report, and the location of the data within the report.}
\begin{tabular}{|l|l|l|}
\hline
\textbf{Metadata} & \textbf{Data from report(s)} & \textbf{Location} \\
\hline
Study Name* & Jahanian 2014 & \\
Reference(s) & \\
\hline
\textbf{Characteristics} & Location in PDF** & \\
Allocation: "randomly assigned" no details reported. & 1985PG30C1P3L7 \\
Blindness: "double blind" no details reported. & 1985PG30C1P3L3 \\
Design: not reported. & \\
Duration: "eight weeks". & 1985PG31C1P3L4 \\
Setting: "Razi Psychiatric Center, Tehran, Iran". & 1985PG30C1P3L5 \\
Participants & Diagnosis: Patients with schizophrenia and tardive dyskinesia (TD) based on DSM-IV-TR diagnosed by psychiatrist. N=40. & 1985PG30C1P3L12-13 \\
Age: range 18-65 years. & 1985PG30C1P3L5 \\
Sex: not reported. & 1985PG30C1P3L17 \\
Outcomes & TD symptoms: no improvement (AIMS). & 1985PG31C1P3L5 \\
Notes & Sponsorship source: "no financial support". & 1985PG33C2P3L1-2 \\
Risk of Bias & \\
Bias & Support Statement from Report & \\
Random sequence generation & "Randomly". No details. & 1985PG30C1P3L7 \\
Allocation concealment & Not reported. & \\
Blinding of participants and personnel & "Double blind". No details. & 1985PG30C1P3L3 \\
Blinding of outcome assessment & "Double blind". No details. & 1985PG30C1P3L3 \\
Incomplete outcome data & Not reported. & \\
Selective reporting & None. Registered protocol is available (IRCT2012092910964N1). & 1985PG30C1P3L1 \\
Other biases & None known. & \\
Outcome & \\
AIMS after Intervention & \\
Rivastigmine & Mean & SD & Placebo & Mean & SD \\
12.5 & 7.0 & 10.3 & 3.1 & 1985PG32T2 \\
\hline
\end{tabular}
\textsuperscript{*} This example study has two references.
\textsuperscript{**} The first 5 digits refer to the file name, PG to pages, C to column, P to page, L to line, and T to Table.
\end{table}
Box 1. Reasons for assessment and concatenation of all references of one study

1. Determining relevance
Screening either references or even an individual full-text report of the references may not provide enough information to convince the reviewer that the study meets the inclusion criteria. Although references are assessed, it is the study that is the unit of analysis.

2. Minimizing the over-counting
Treating data from different references of the same study as separate studies will result in over-counting and spurious results.

3. Minimizing the under-counting
Using data from only a selection of references of the study may fail to identify important outcomes in other relevant references of that same study. For example protocols of studies may report use of many more outcome measures than are finally reported in any one reference.

Table 2. Levels of specialized registers

Beginners’ level or reference-based register

a. Setting the scope of register
b. Developing, running, documenting, and saving the search strategies
c. De-duplicating and curate the search results
d. Screening the search results
e. Developing minimum dataset for each reference record
f. Importing the reference records
g. Maintaining the reference records
h. Locating the full texts of all reference records

Intermediate level or study-based register

a. Developing coding scope and guideline for studies
b. Coding or extracting general data from each reference
c. Concatenation, merging, and cleaning the study records
d. Maintaining and updating the study records

Advanced level or automated study-based register

a. Classifying the studies under each review title ready to be done
b. Developing a machine-readable dataset for each study
c. Extracting all data from each study

d. Maintaining and updating the study records

...
malleable design to be open to future developments. Much has already been documented regarding such efforts for projects such as Distiller SR, Covidence, and Systematic Review Data Repository (SRDR).

**Challenges**

1. **The concept**

The currency of the Information Specialist in health care has, for so long, been the individual reference (1 record for 1 reference) and making the jump to considering studies (1 record to potentially many references of 1 study) as the primary unit of information can be a conceptual challenge. After all, even cursory use of databases such as MEDLINE can cause nagging discomfort that information is being identified for reviewers that is less ordered and sorted than is ideal. For example, searching for CATIE, a large trial acronym, in the title field of MEDLINE will reveal multiple references (170 references at the time of writing). More latterly, solutions to this discomfort have been conceived and so the issue is more acknowledged.\(^{20,29-31}\) The process of systematic reviewing does help make this conceptual jump with the use of ‘study tags’ under which all relevant references are listed.

In a physical library with books, a librarian undertakes collection development, cataloging, classification and dissemination of information. A reference-based register with just collection development and no coding is the equivalent of maintaining a stack of books with no classification. Whilst ‘save the time of the reader’ is one of Five Laws of Library Science\(^{41}\) ‘save the time of the reviewer’ is, we argue, best addressed in a study-based register. Also having a reference-based register ‘might work’ for some research groups, however, study-based register ‘might work even better’ saving the time and money for both the Information Specialists and systematic reviewers.

2. **Responsibility**

In reality, the process of identification of studies is a shared responsibility between Information Specialist and reviewer. Each has different skills to bring to the process. The former has competency in the identification of records, and knowledge of what each record should contain. For example, with reference records indexed with study identifiers (e.g., trial registry number), the Information Specialist is in a pivotal position to help the reviewer avoid needless effort linking references to a study. The reviewer, however, having inspected the detail of each reference, should be able to supply an authoritative study record back to the Information Specialist for their use or for the next reviewer.

Even now the responsibility for concatenation is shared. However, study-based registers allow this responsibility to be undertaken more easily by the person who maintains the register. The nature of systematic reviewing is that everything is double checked. By an iterative process, the study record evolves to be an increasingly accurate report of the primary investigation. To continually pass this responsibility down the line to reviewers creates an unnecessary waste and opportunity for inaccuracy.

3. **Practicalities**

For the gains which we outline above, development of such a study-based register involves an investment of effort – some pain - often from Information Specialists. Currently, Information Specialists working in systematic reviewing of health care interventions are busy and the thought of further work and/or responsibility may be unwelcome. However, much current work is inefficient and we argue the work creating a study-based register is an investment. The ‘pain’ involved makes the role of Information Specialist much more sophisticated and prepares that person to manage a register suitable for the data needs of the 21st century.

Coding of study happens by a shared iterative process as outlined above. Many reference records already contain study codes which can be imported into the study record. For example, the PT field of MEDLINE may contain “Randomized Controlled Trial” – a methodological term relating to the study, or the SI field contains the International Standard Randomized Controlled Trial Number (ISRCTN) – again data that relates to the study rather than the reference. These data can be imported automatically into the study record at no cost of time. As reviewers ‘use’ the study, adding to the complexity but also the utility of the record, more sophisticated data from this investment of effort can be curated and stored ready for the next reviewer.

4. **Software**

Although there are online clinical trials register such as ClinicalTrials.Gov, ISRCTN, and WHO International Clinical Trials Registry Platform (including 16 trial registers), however, none of these are study-based registers and none are aimed to support systematic reviews and provide a very limited number of fields for each study record. Even before the initiation of these sources, in 1995, Cochrane Stroke Group started using study-based register and mental health followed.\(^{44-49}\) A year after Stroke Group, UK Cochrane Centre supported MeerKat Working Group to develop a study-based register system.\(^{21,50}\) In 2003, there were 10 Cochrane groups using MeerKat and 5 other groups considering its use.\(^{21}\) By 2005, at least 12 Cochrane groups (out of 37 responses) were maintaining a study-based register on MeerKat, ProCite, Reference Manager or RefTrak.\(^{25,51-53}\) In 2008, Cochrane started developing a new program, Cochrane Register of Studies (CRS), to be used by all groups.\(^{54-59}\) A survey in 2014 showed that 8 out of 29 respondent groups are using a study-based register to some extent.\(^{12}\)

There are several reference management programs and some study-based register programs. Despite the
decision of Cochrane to move to CRS as the only program for managing information on both references and studies, groups do tend to use other programs as well as CRS. Although others have pioneered software for study-based registers, there are few packages that are generic, accessible, or customizable. EndNote, ProCite, and Reference Manager are popular reference management programs that can be modified to include some features of study-based registers. These bibliographic packages do not lend themselves easily to this adaptation and relational databases have considerable advantages. Since the purpose of programs such as EndNote or Reference Manager is not primarily managing studies for systematic reviews, using them in a study-based fashion requires time and effort – and there is an element of fitting a ‘square-peg-in-the-round-hole’. Tailor-made relational programs such as MeerKat or RefTrak are much better and potentially more malleable. CRS continues to evolve.

No matter what software or application is used, the transition to a study-based register necessitates the acquisition of a skill-set to use the package to full capacity from the study-perspective. This transition is helped by practice, training and mentoring.

5. Data ownership
As mentioned before, there are other resources, such as online registries of trials, that do provide some useful data for studies; however, there are certain barriers in terms of copyright and legal issues such as ‘who owns the data’. Such limitations make it hard to use and share the data openly and import or share them from other resources in study-based registers.

6. Study designs
Since the current development of study-based registers in Cochrane, is mainly focused on RCTs/CCTs, and because we work within that organization, we have not discussed involving other study designs. There is, however, a possibility to include all empirical study designs in study-based registers of the future. This may cause new challenges for the development of the registers involving different sets of meta-data. Some Cochrane groups are already considering involving more diverse study designs in their reviews. However, these other study designs are not a priority for consideration in the technological development of specialized registers. The development of such registers has made it possible to link the studies from pre-clinical sciences to clinical sciences. Such link could reduce the waste by avoiding a clinical trial research where a systematic review of pre-clinical studies about ineffectiveness of an intervention exists.

Gains
Having a study-based register gives review groups the advantage of knowing with some accuracy how many systematic reviews (or at least comparisons) there are to cover in a topic area – and to more accurately estimate the future work-load. Such knowledge allows accurate and efficient prioritization of effort.

Current resources, if used effectively, are sufficient to increase productivity. Using the study as the currency of the Information Specialist, modifying that role within a review group to create and maintain the study-based register, and finally using such a register will greatly increase the pace and efficiently of information exchange.

Information Specialists could then undertake regular – perhaps semi-automated – searches, screen, and code and include relevant records within the study-based register. New references of existing studies would be added to the existing record and, if this study is used within an existing review, the reviewers would be notified. New studies relevant to existing review topics would also cause reviewers to be alerted. Rather than passively waiting for reviewers to request update searches Information Specialists would be proactive in helping the update. From the ordered study-based register studies can be linked to topics and be instantly ready for new reviewers. Should data from a study have been extracted for use in an existing review, these detailed data could be supplied from a sophisticated study-based register in an appropriate format to anyone undertaking a new review necessitating the use of the same study.

Conclusions
The ultimate goal of having a study-based register is to facilitate efficient production of systematic reviews
providing rapid but accurate evidence for decision-makers. The future will involve increasing automation. Document optimization now allows much more reliable auto-data extraction.79 Programs already exist for semi-automated data extraction from randomized trials5 and text mining techniques are increasingly sophisticated. The automatic synthesis of data is not far away, perhaps driven by users’ needs rather than those of policymakers. Limited automatic write-up of synthesized evidence already exists.79–81 These next years will see a swift synthesis of best and personalized evidence of the effects of health care in the hands of anyone. At the heart of this exciting prospect should be the role of the Information Specialist – but a role fit for the 21st century and not one that is dated and wasteful. Moving from reference-based register toward study-based register is, we think, inevitable. The infrastructures are ready for such movement.

Ethical issues
There is none to be declared.

Competing interests
No competing interests to be disclosed.

Acknowledgements
We would like to thank Judy Wright for providing available documents on study-based registers for us and also Gail Higgins for her invaluable comments on the final draft of the manuscript.

Supplementary Materials
Supplementary file 1 contains a Randomized Controlled Trial Study in XML Format.

References
2. Gonzalez A. PNA/08-02: “Development and implementation of a methodology to discover knowledge in corporative database as support to the organizational making decision process”. CUBAEnergia; 2002.
17. CRS Team. Studdication: maintaining your study based register. CRS Webinar; 2013.
38. Noel-Storr A. Standardisation: study-based meta data. CRS Webinar; 2015.
50. MeerKat Working Group, editor. MeerKat:getting the most out of your study-based Specialized Register. Melbourne, Australia: Cochrane Qolloquium; 2005.
53. Wright J. Setting up and maintaining a study-based register. Cochrane Schizophrenia Group; 2006.
55. Foxlee R, editor. The development of the new Cochrane Register of Controlled Studies. Freiburg, Germany: Cochrane Qolloquium; 2008.
73. Noel-Storr A, Malouf R, editors. Launch of ALOIS - an online study-based register of research from the Cochrane group. 4th UK Dementia Congress; 2009; Harrogate, UK.
78. McClennaghan E, Roberts S, Adams CE. So – how many reviews do you have to do? Cochrane Schizophrenia Group; 2012.