

# Developing a Registry of Patient Registries: Options for the Agency for Healthcare Research and Quality (Draft Report)

---

Task Order No. 3  
Contract No. HHSA290200500351

***Report Prepared by:***

The Outcome DEcIDE Center  
201 Broadway  
Cambridge, MA 02139-1955

Draft Submitted May 31, 2009

## Table of Contents

EXECUTIVE SUMMARY .....	3
INTRODUCTION .....	5
Project Objectives .....	6
Rationale for a Registry of Patient Registries .....	6
Objectives of a Registry of Patient Registries .....	9
ANALYSIS OF EXISTING RESEARCH REGISTRIES.....	11
Identification of Registries .....	12
Description of Existing Databases.....	12
Clinicaltrials.gov .....	12
DoCDat .....	14
HSRProj .....	16
PDQ Cancer Clinical Trials Registry .....	17
World Health Organization (WHO) International Clinical Trial Registry Platform (ICTRP) .....	18
Comparison of Existing Databases .....	20
Relevance of Existing Databases for a Registry of Patient Registries .....	21
PROPOSALS FOR REGISTRY OF PATIENT REGISTRIES .....	23
Should There Be a Registry of Patient Registries? .....	23
What Are the Requirements for the Registry of Patient Registries? .....	23
What Approaches Might AHRQ Pursue to Develop and Maintain a Registry of Patient Registries? .....	24
Proposal 1: Develop the Registry of Patient Registries <i>De Novo</i> .....	24
Proposal 2: Collaborate with Clinicaltrials.gov to Build an Integrated Patient Registry Module .....	25
Proposal 3: Develop Hybrid Approach that Leverages Capabilities of Clinicaltrials.gov & AHRQ .....	26
What Would Motivate Researchers to List Their Registries in the Registry of Patient Registries?.....	26
What Are the Costs of Setting Up and Maintaining a Registry of Patient Registries?.....	27
CONCLUSIONS.....	28
REFERENCES .....	30

## **EXECUTIVE SUMMARY**

The purpose of this project is to examine the potential value and feasibility of listing patient registries in a research registry and to explore options for developing a registry of patient registries. The goal of this paper is to provide actionable information to the Agency for Healthcare Research and Quality (AHRQ) for developing a registry of patient registries that will be relevant to the needs of the Medicare, Medicaid, and other Federal health care programs and will reflect the overall goals of the Effective Health Care program. While many of the findings and recommendations in this paper may also be true for other observational study types, the scope of this effort was specific and limited to patient registries.

To explore the value of and identify objectives for a registry of patient registries, interviews were conducted with stakeholders from the public and private sectors, including the U.S. Food and Drug Administration, the Centers for Medicare and Medicaid Services, institutes of the National Institutes of Health, professional associations, medical societies, and foundations. These interviews identified several potential benefits of a registry of patient registries, as well as some concerns.

The interviews also led to the development of several potential objectives for a registry of patient registries. These include: 1) to provide a searchable central listing of patient registries in the United States to enable interested parties to identify registries in a particular area of interest (to promote collaboration, reduce redundancy, and improve transparency); 2) to encourage and facilitate the use of common data elements and definitions in similar conditions (to improve opportunities for sharing, comparing and linkage) through the listing of and searching on such elements; 3) to provide a central repository of searchable summary results (including for registries that do not have results published in peer-reviewed literature); 4) to offer researchers a search tool to locate existing data sources (from either ongoing registries or closed registries) to request for use in new studies (secondary analyses, linkage studies); and 5) to serve as a recruitment tool for both providers and patients.

In terms of functional requirements, stakeholders identified several key features of a registry of patient registries. First, it would contain a list of existing patient registries. For each included registry, information on the registry purpose, study design, disease area, patient population, participating sites, enrollment target, status of enrollment, geographic location, and data collection would be available in a standard format. Second, the web-based system would enable visitors to search for patient registries using a range of data elements (e.g., study design, disease area, geographic location, etc.). The search engine would also use the MeSH hierarchy of terms to help guide users to appropriate studies. Third, summaries of registry analyses could be available in standard table formats. Lastly, the website could include a description of data assets, collaborative uses of interest to the data asset owner, and contact information for further data use discussions. As an extension of this, a secure repository (and appropriate access rules) might be developed to house registry databases donated by patient registries that have closed.

A review of existing databases of patient registries and other clinical studies confirmed that the existing databases do not meet the objectives or functional requirements identified by the stakeholders. The databases that were reviewed for this project included Clinicaltrials.gov, DoCDat, Health Services

Research Projects in Progress (HSRProj), PDQ Cancer Clinical Trials Registry, and the WHO International Clinical Trial Registry Platform. The report provides detailed information on each database.

Because the existing databases do not meet the needs of stakeholders, it was necessary to consider potential approaches for an AHRQ-sponsored registry of patient registries. This report examines the strengths and limitations of three options: 1) develop the registry of patient registries *de novo*; 2) collaborate with Clinicaltrials.gov to build an integrated patient registry module; and 3) develop a hybrid approach that leverages the capabilities of Clinicaltrials.gov and AHRQ. The report describes the importance of a means to minimize duplication of effort for investigators who seek to have information reside in both a registry of patient registries and Clinicaltrials.gov. The report also discusses possible ways to incentivize registry sponsors to list their registries in the database as well as a more active outreach approach as an alternative.

The report concludes that there is a clear need among stakeholders for a registry of patient registries, but the lack of incentives or pressures for registry owners to participate is a critical barrier to a successful program. Given the participation risks, the most practical approach may be one that begins with the core functionality of listing and search and builds other components over time. Additional interviews and discussions with both stakeholders and registry owners are highly recommended as a next step to help AHRQ further define requirements, priorities, incentive structures, and funding options.

## INTRODUCTION

Within the United States (U.S.), a large number of clinical research projects are underway at any given time. These range from controlled studies of experimental products to observational studies of approved, marketed products to retrospective studies using electronic health records or claims data. Due to the tremendous amount of resources in terms of time, money, knowledge, and willing patients devoted to this research, it is in the best interest of the public to ensure that the results of such research are disseminated and that research is not duplicated unnecessarily. In addition, some patients may consent to participate in medical research with the understanding that their efforts will result in increased knowledge in the medical community. Dissemination of the study results is necessary from an ethical standpoint to fulfill this promise to the study participants.

Despite the clear public benefit from reporting of clinical research results, some reviews have found that over one third of clinical trials never publish their results in medical literature.<sup>1</sup> As part of an effort to increase transparency in clinical research, the U.S. Food and Drug Administration (FDA) Modernization Act of 1997 provided for the creation of the Clinicaltrials.gov registry. All efficacy drug trials conducted under Investigational New Drug applications for “serious or life-threatening diseases and conditions” are required to be included in this registry.<sup>2</sup> Currently, the registry includes over 73,000 trials funded by the Federal government and private industry. The trials are being conducted in all 50 states in the U.S. and in 167 countries.<sup>3</sup>

The Food and Drug Administration Amendments Act of 2007 (FDAAA) expanded the types of trials that must be listed in the registry to include any controlled clinical investigation, other than Phase 1 investigations, of a product that is subject to FDA regulations, and any controlled trials with health outcomes of devices subject to FDA regulation.<sup>4</sup> The law also increased the amount of information that must be included for each trial and mandates the submission of results data. Clinicaltrials.gov is in the process of implementing these recommendations.

In addition to Federal legislation mandating that studies be included in Clinicaltrials.gov, researchers who wish to publish the results of their studies in any major medical journals are required to list the studies in a research registry. The 2005 Statement from the International Committee of Medical Journal Editors (ICMJE) requires that any prospective trial be registered as a condition of publication.<sup>5</sup> The ICMJE policy identifies suitable registries, including Clinicaltrials.gov, International Standard Randomized Controlled Trial Number (ISRCTN), Australian Clinical Trials Registry, Netherlands Trial Registry, Japanese Ministry of Education (UMIN) Clinical Trials Registry,<sup>6</sup> and any of the primary registries that participate in the World Health Organization’s (WHO) International Clinical Trial Registry Platform (ICTRP).<sup>7</sup>

Clinicaltrials.gov and the accompanying legislation, as well as the requirements of the ICMJE, are important steps forward in increasing transparency in clinical research and improving access to the results of clinical research. However, these requirements only apply to controlled investigations. A significant amount of clinical research is uncontrolled, or observational. For example, prospective patient registries, prospective cohort studies, retrospective case-control studies, and retrospective analyses of electronic health data (e.g., claims data, electronic health record data) are all uncontrolled

studies. These types of studies are not required, by either U.S. law or the policy of the ICMJE, to be listed in Clinicaltrials.gov or any other research registry.<sup>8</sup>

Observational studies are not prohibited from being included in Clinicaltrials.gov, but the researchers lack incentives to list their studies in the registry. As a result, of the 73,000 trials listed in Clinicaltrials.gov, only 11,000 are described as observational, and only 7,000 are described as prospective observational studies. This raises four questions. First, is listing observational studies in a research registry a good use of resources? Second, if it is a good use of resources, where should observational studies be listed (i.e., in an existing registry or in a new registry)? Third, if such a listing registry existed, what components of prospective observational studies in particular should be listed and why? Finally, what would motivate researchers to list their observational studies in a research registry (i.e., incentives)?

### **Project Objectives**

The purpose of this project is to examine the potential value and feasibility of listing patient registries in a research registry and to explore options for developing a registry of patient registries. The goal of this paper is to provide actionable information to the Agency for Healthcare Research and Quality (AHRQ) for developing a registry of patient registries that will be relevant to the needs of the Medicare, Medicaid, and other Federal health care programs and will reflect the overall goals of the Effective Health Care program. While many of the findings and recommendations in this paper may also be true for other observational study types, the scope of this effort was specific and limited to patient registries.

The paper begins by establishing the rationale for a registry of patient registries and explaining the objectives of such a registry based on input from stakeholders. Next, the paper reviews existing research registries to determine if any of these meet the objectives of a registry of patient registries. Lastly, the paper makes recommendations for moving forward with a registry of patient registries and explores three proposals or approaches for such a project. Each proposal includes information on how registries would be identified and chosen for inclusion; what incentives would be available for patient registries to submit information; what information would be included; how information would be obtained and verified; and the costs for different proposals.

The paper also identifies limitations and weaknesses of the current offerings and describes potential justifications for considering a registry of patient registries. The information in this paper is based on literature reviews, interviews with stakeholders, and an analysis of existing research registries.

### **Rationale for a Registry of Patient Registries**

In order to understand the potential rationale for a registry of patient registries, it is useful to compare and contrast the drivers of such a listing registry to the drivers that led to the creation of listing registries for clinical trials. Starting in the early 1990s, progressive interest in developing and ultimately mandating participation in clinical trials registries was largely driven by the need to increase the transparency of clinical research, as it was perceived that unfavorable results were being systematically underreported.<sup>9</sup> Listing trials, it is believed, increases the likelihood that the results of medical experiments will be made available for the public good—an ethical obligation given that human subjects

voluntarily participate in research largely for this reason. The natural extension of listing clinical trials is therefore the listing of results, both in the form of publications and in the form of summary results for trials that are not ultimately published in peer reviewed journals. In the case of trials with negative results, such information might prevent future similar trials from being developed. An overarching goal of such efforts is increasing public trust of and potential participation in clinical research, a benefit that accrues to the medical research establishment as a whole. The arguments against a registry of trials included the potential disclosure of confidential information and an inefficient use of resources. However, in the case of trials, the ethical imperative outweighed these arguments. In 2000, Clinicaltrials.gov was launched in the U.S. and subsequently endorsed by the ICMJE. Since that time, other mandates, such as the Food and Drug Administration Amendments Act of 2007, have further extended requirements for reporting to Clinicaltrials.gov and for using trial identifiers that enable cross-linking between databases to occur. As described further below, these actions have successfully led to large increases in the listing of relevant clinical trials in the U.S. on Clinicaltrials.gov and similar programs in other countries. For example, the National Institutes of Health reported that clinical trial registrations increased from approximately 250 per week to 440 per week between December 13, 2007 and January 12, 2008.

A patient registry is defined as “an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes.”<sup>10</sup> Common purposes for patient registries include evaluating the safety, effectiveness, or quality of medical treatments, products, and services and studying the natural history of diseases. Some registries are developed and maintained solely to assist in care delivery, coordination, and quality improvement, but many serve broader research purposes. While there may be less of an ethical imperative to list non-experimental studies like patient registries in a research database, some would argue that not unlike a clinical trial, registries with a research purpose do benefit from and impose some burdens (e.g., time to complete surveys, risk of loss of privacy) on their participants. Therefore, patient registries have obligations similar to clinical trials to ensure that the knowledge gained from these efforts is generalized for the public good.

There are a number of reasons why establishing a registry of patient registries (for patient registries with a research purpose) would be highly beneficial to a number of stakeholders and consistent with the research purposes of those registries. First, given the finite amount of resources that can be devoted to clinical research, reduction of unnecessary duplication of effort should be a high priority. Listing patient registries would allow researchers who are considering a new registry to identify similar studies and avoid unnecessary duplication of research questions or populations. Enough detail would be required to enable a sophisticated registry developer to understand if the methodologies applied were sufficiently rigorous for the particular purpose in mind. A contact point would be beneficial to encourage a direct conversation between registry developers. For example, researchers undertaking a patient registry of cardiovascular disease among women under the age of 55 in a certain region may find several similar registries underway with which they might collaborate or choose to differentiate to better answer certain questions. In either case, overall resource utilization would be improved.

Second, the sharing of data element lists through a listing registry would promote standardization and increase the likelihood that future registry datasets in the same conditions could be later compared, joined, or linked. Third, there is considerable interest and opportunity in linking registries and other datasets together to construct a longitudinal picture of disease and to obtain additional outcomes (e.g., longer term or a different type, such as costs). Knowing what registries exist, either actively in operations or as archived registry databases, would be a highly valuable resource for investigators to connect on possible collaborative linkages. In addition to linkages, many researchers would be interested in opportunities to identify registries to request specific analyses or to request data to perform additional investigations. The presence of such a resource would be a great boon to observational research in general.

The provision of summary results, even for registries that did not publish their results, will likely provide some useful information on the populations that were studied. If a results database were also available, certain registries might choose to archive their data in the database for future investigations if the purpose or funding for their own registry expired. Lastly, a listing of patient registries could serve as a recruitment tool for these registries. Both providers and patients might seek to participate in a registry if identifiable through such a database. These are just a few potential uses for a registry of patient registries. While some of the uses overlap with those for which clinical trials registries exist, some are unique to registries and other observational studies.

At the same time, depending on the nature of the incentives, a registry of patient registries may raise several concerns. Registry owners may have proprietary claims on data elements, definitions, and data that may significantly limit the degree to which such information can be put into the public domain. Certain registries that do not have an initial research purpose (e.g., those designed for quality improvement) might not list themselves in the database initially, but may wish to do so later when they realize that they have collected worthwhile information to share. Were ICMJE to expand their current rules to include registries, these registries would be barred from publication because of the failure to register initially. Another concern relates to patient privacy. While great care is taken today in linking datasets to avoid re-identification, this analysis is done for each linkage exercise. What is less clear is what the risk is for re-identification if a series of registry linkage results with different databases were placed in the public domain. Another issue is the potential burden of participating. The need to post and update information can create significant burden on the registry owner, depending on the level of detail to be maintained. For the organizations already voluntarily registering their patient registries on Clinicaltrials.gov, there would also be significant resistance to any duplication of effort required to maintain data in both Clinicaltrials.gov and a new registry of patient registries database should they not be fully interoperable.

The above discussion outlines several clear benefits as well as concerns regarding a potential registry of patient registries. How such a resource is developed, governed, and operated, and how registry owners are incentivized to participate ultimately will determine if the net benefits outweigh the concerns.

## Objectives of a Registry of Patient Registries

Recognizing that a registry of patient registries might serve multiple goals for different stakeholders, registry owners and potential investigators, the next step is to determine what specific goals an AHRQ-supported registry of patient registries should serve.

Stakeholders from the public and private sectors were asked whether a registry of patient registries would provide value, and, if so, in what ways. Government stakeholders included the Coverage and Analysis Group of the Centers for Medicare and Medicaid Services (CMS); the Food and Drug Administration (FDA); the National Institutes of Health (National Cancer Institute (NCI) and National Heart Lung and Blood Institute (NHLBI)), and the National Center on Birth Defects (NCBD). Private sector stakeholders included medical professional societies, such as the American College of Rheumatology (ACR), American College of Chest Physicians (ACCP), and the American Academy of Neurology (AAN), and private foundations, such as the Cystic Fibrosis Foundation (CFF). Also interviewed was AcademyHealth, an organization of health services researchers that represents other potential secondary users of data from registries.

Federal stakeholders (CMS, FDA) emphasized the value that a registry of patient registries would bring as a means to identify existing registries and as a potential tool for encouraging standardization of data elements. Identification of existing patient registries would enable these stakeholders to more easily contact and potentially leverage these data sources for other projects. The FDA specifically cited the Sentinel Initiative and other safety monitoring programs as examples.

NCI, NHLBI, and NCBD all expressed interest in the possibility of identifying registries where there are data sharing capabilities in place or where the registry sponsors are looking for collaboration opportunities. It was noted, for example, that there are many long-term cohort studies conducted within the National Institutes of Health (NIH) that may be useful for extramural research; however, these studies may be difficult for non-NIH researchers to identify. NHLBI suggested that a registry of patient registries might also help existing registries recruit new, and particularly younger, less experienced, investigators. In terms of barriers, it was pointed out that researchers may be reluctant to devote resources to registering their study without a mandate or strong incentives. Despite the value, some questioned whether such an effort would be an appropriate use of resources for AHRQ and for registry sponsors.

In general, the medical professional societies interviewed (ACR, ACCP, AAN) viewed a registry of patient registries as a useful resource. It was felt to be very important that any submitted information be carefully vetted for accuracy and should include information on the quality, strengths, and limitations of the included registries. The ACR stated that the information available from such a registry of patient registries could be used to help build collaborative networks among registries and increase opportunities to share and compare data. All of the societies noted that specific information on the types of data collected would be particularly important for identifying potential collaborators.

Most societies stated that they would be interested in listing their registries, if the process was not overly burdensome. There was general agreement that there would need to be some sort of incentive

to do so, whether it is required for publication, mandated by Federal law, or necessary for some other reason. Based on the conversations with societies, the primary barrier to listing registries is the burden of listing the registry and keeping the information up to date. The CFF manages a long-standing registry with very high national participation by both providers and patients. This group provided the perspective of a private, patient-centric foundation. CFF stated that they receive many requests from researchers for access to data and inquiries from others interested in setting up a registry. They suggested that a registry of patient registries would have value for people who are running registries so they can connect with other registry sponsors and perhaps collaborate or learn from the other registry. The CFF noted that they themselves would be interested in listing their registry in such a database.

AcademyHealth offered insights into how the registry of patient registries may have value for health services researchers. Similar to other groups, they noted the need for more transparency in the field of patient registries to encourage collaboration among registries and to improve the possibility for linking data sets or conducting research using existing data sets. The idea of including summary results was particularly attractive to this group. They also suggested including features of Clinicaltrials.gov, such as the indexing with National Library of Medicine (NLM) Medical Subject Headings (MeSH) and providing the tabular view of results to facilitate downloading the results. In terms of barriers, AcademyHealth agreed with other stakeholders that the biggest barrier would be finding ways to incentivize registry sponsors to list their projects and update the information.

Based on these discussions with stakeholders, several broad potential objectives for a registry of patient registries were identified:

1. To provide a searchable central listing of patient registries in the U.S. to enable interested parties to identify registries in a particular area (to promote collaboration, reduce redundancy, and improve transparency).
2. To encourage and facilitate the use of common data elements and definitions in similar conditions (to improve opportunities for sharing, comparing, and linkage) through the listing of and searching on such elements.
3. To provide a central repository of searchable summary results (including results for registries that have not published their findings in peer-reviewed literature).
4. To offer researchers a search tool to locate existing data (from either ongoing studies or closed studies) to request for use in new studies (secondary analyses, linkage studies). This was primarily envisioned as a description of data assets and contact information to facilitate collaborative discussions between parties. However, it was mentioned that expiring registries might occasionally donate their databases to a repository for public use.
5. To serve as a recruitment tool for researchers and patients interested in participating in patient registries.

In addition, we asked stakeholders to identify the key functions that a registry of patient registries should be able to support. These are listed in Table 1.

***Table 1: Potential Functional Requirements of a Registry of Patient Registries***

Objective	Functional Requirements	Priority Level
<b>List existing patient registries</b>	Allow registered users to enter registries and update information on registries. For each registry, list: <ul style="list-style-type: none"> <li>• Registry purpose</li> <li>• Condition/intervention</li> <li>• Inclusion/exclusion criteria</li> <li>• Planned number of participating sites</li> <li>• Patient enrollment target</li> <li>• Status of enrollment (i.e., actively recruiting participants?)</li> <li>• Geographic location</li> <li>• Study design</li> <li>• Description of data collection, including data elements, use of validated instruments, and follow-up duration</li> </ul>	High – this objective was mentioned by all stakeholders as a key function of a registry of patient registries
<b>Enable users to search for registries</b>	Allow users to search for registries by any of the data elements identified in the “List existing patient registries” section. Provide sophisticated search engine that uses Medical Subject Headings (MeSH) from PubMed, builds on a hierarchy of terms results, and corrects misspellings.	High – this objective was mentioned by all stakeholders
<b>Summarize registry results</b>	For each completed analysis, provide: <ul style="list-style-type: none"> <li>• Population studied</li> <li>• Inclusion/exclusion criteria</li> <li>• Analytical methods</li> <li>• Patient demographics summary</li> <li>• Summary of analysis findings and interpretation of findings</li> <li>• Publication citations</li> </ul>	Low – this objective was only mentioned by one stakeholder
<b>Provide repository for registry databases</b>	For each registry, provide: <ul style="list-style-type: none"> <li>• Information noted in the “List existing patient registries” section</li> <li>• Description of data assets</li> <li>• Time frame for data collection</li> <li>• Collaborative uses of interest to the registry developer (e.g., what is permitted under the authorizations received)</li> <li>• Contact information for data use discussions</li> </ul>	Medium – this objective was mentioned by several stakeholders, but not all.

## ANALYSIS OF EXISTING RESEARCH REGISTRIES

With the justification for a registry of patient registries established and the objectives determined, the next step is to identify and examine existing research registries. The purpose of this analysis is to

determine if, and how, existing systems address the objectives stated above (either generically in trial registries, or specifically in observational research registries to the extent that they exist).

## Identification of Registries

Existing research registries were identified using multiple sources. First, the registries that are considered acceptable under the ICMJE policy were included.<sup>7</sup> Registries also were identified using Internet search engines, such as Google and Google Scholar.<sup>11,12</sup> To identify relevant published literature, the team used PubMed, a search service of the National Library of Medicine and the National Institutes of Health.<sup>13</sup> Using this search tool enabled the team to identify research registries cited in published literature. Research registries mentioned in interviews with stakeholders were also included.

This process yielded the following list of research registries:

- Clinicaltrials.gov
- DoCDat
- Health Services Research Projects in Progress (HSRProj)
- PDQ Cancer Clinical Trials Registry
- WHO International Clinical Trial Registry Platform
- Australian Clinical Trials Registry
- Netherlands Trial Registry
- UMIN Clinical Trials Registry (Japan)

The Australian Clinical Trials Registry, the Netherlands Trial Registry, and the UMIN Clinical Trials Registry were excluded from further analysis since these registries are country-specific. More information on the remaining five research registries is included below.

## Description of Existing Databases

### Clinicaltrials.gov

Clinicaltrials.gov provides information on Federally and privately supported clinical trials conducted in the U.S. and internationally. Trials and observational studies in a wide range of disease areas are included in the registry. The purpose of the site is to provide easy access to trial information to patients, family members, health care professionals, and other members of the public. The database currently includes over 73,000 trials.

As noted above, Clinicaltrials.gov was developed in response to the Food and Drug Administration Modernization Act of 1997. The site, the result of a collaboration between the National Library of Medicine (a branch of the National Institutes of Health) and the FDA, launched in February of 2000. Registration of controlled clinical studies is required by law in the U.S. In addition, the ICMJE requires registration of interventional studies as a condition for publication. Clinicaltrials.gov is an acceptable registry to meet both the U.S. legal requirements and the policy of ICMJE. The scope of Clinicaltrials.gov was expanded with the passage of the Food and Drug Administration Amendments Act of 2007 (FDAAA). FDAAA requires the registration of certain trial results and expands the types of trials included in the

registry. A major update to the Clinicaltrials.gov website has been necessary to meet these new requirements.

For each listed study, Clinicaltrials.gov includes the following data elements:

**Table 2: Clinicaltrials.gov data elements**

Category	Data Elements
Abstracts of clinical study protocols	<ul style="list-style-type: none"> <li>• Summary of the purpose of the study</li> <li>• Recruiting status</li> <li>• Inclusion/exclusion criteria for patient participation</li> <li>• Location of the trial</li> <li>• Specific contact information</li> </ul>
Additional information that may help a patient decide whether to consider a particular trial:	<ul style="list-style-type: none"> <li>• Description of study design</li> <li>• Phase of the trial</li> <li>• Disease, condition, drug, or therapy under study</li> </ul>
Additional Links	Links to health information (e.g., MEDLINEplus, PubMed) that may help to place the trial in the context of a patient's overall medical care

Study sponsors submit information on their studies using a web-based tool, the Protocol Registration System, in accordance with FDA guidance.<sup>14</sup> Sponsors must submit information on new studies no later than 21 days after enrollment for the trial begins. Studies will remain in the registry indefinitely, as it is intended to be both a registry of open and closed trials.<sup>15</sup> Both Clinicaltrials.gov and the FDA utilize the same unique trial identifier to allow linkage. Study sponsors are responsible for keeping information on their study up to date; new information is reviewed by the National Library of Medicine before being posted to the site.

The results database, which is still being refined, will allow posting of basic data tables that were mandated in the FDAAA law. The key information to be posted includes a table of baseline demographics information; a "CONSORT" style diagram; and tables of outcomes.

A major component of Clinicaltrials.gov is its search capability. One of the advantages of Clinicaltrials.gov is that it is able to leverage other resources under the National Library of Medicine umbrella such as PubMed. Clinicaltrials.gov data is available via download in XML format and can readily be tied to other websites, such as the National Cancer Institute's PDQ Cancer Trials Registry (see below for more information). The registry makes significant efforts to minimize duplicate study listings and to provide a search engine that identifies trials meeting a user's specifications. The search functionality uses MeSH terms from PubMed to build on a hierarchy of terms, such that a search for one term will return results for related terms as well. For example, a search for "hair loss" will return results for "hair loss" as well as "alopecia." The site also uses an automatic spell check tool to assist users who are uncertain of the correct spelling of a word.

Clinicaltrials.gov has relatively complete listing information from more than 73,000 clinical studies. An evaluation of the completeness of the results database is not yet feasible.

### DoCDat

The DoCDat database, based in the United Kingdom (U.K.), is intended to serve as a resource for locating databases with patient level and administrative data on clinical care in the U.K. The database is funded, maintained, and developed by the Information Centre for Health and Social Care, which is part of the National Health Service. The primary audiences for the database are those involved in clinical governance, health services management, and health services research. The objectives of DoCDat are to use a standardized method to assess the quality (in terms of completeness and accuracy) of clinical databases and to establish a web-based directory of databases the U.K.<sup>16</sup> The purpose of the project is to facilitate the use of existing clinical databases, avoid duplications of effort, and improve database quality.

For each listed clinical database, DoCDat includes a significant amount of information regarding the database that should allow any potential user of the data to determine their interest in contacting the database owner. DoCDat goes beyond simply listing general descriptions and contact information. In addition to a description of the reference population, it includes a full description of data elements, periodicity of data collection, data quality assessments, whether the subjects are informed of data collection or consented for participation, and what standard data elements for data linkage are utilized. Simple lists of questions to rate data quality are also completed. The table below summarizes the information that is collected:

**Table 3: DoCDat data elements**

Category	Data Elements
Background	<ul style="list-style-type: none"> <li>• Date last updated</li> <li>• General description of database</li> </ul>
Reference population	<ul style="list-style-type: none"> <li>• Inclusion criteria</li> <li>• Whether database is longitudinal</li> <li>• Geographic area</li> <li>• Time frame covered by data</li> <li>• Time frame for sampling (e.g., continuous, periodic, etc.)</li> </ul>
Data set content	<ul style="list-style-type: none"> <li>• Number of individuals or episodes of care</li> <li>• Data collection questionnaire(s)</li> </ul>
Data linkage	<ul style="list-style-type: none"> <li>• Are nationally approved codes used for identifying subject, clinician, or institution?</li> <li>• Are other databases linked to this database?</li> </ul>
Security	<ul style="list-style-type: none"> <li>• Where are electronic data held?</li> <li>• Where is a back-up version of the data held?</li> <li>• Are paper forms stored?</li> </ul>
Confidentiality	<ul style="list-style-type: none"> <li>• How are records stored (identifiable, reversibly anonymised, irreversibly anonymised)?</li> <li>• Are subjects informed that data are collected and used?</li> <li>• Have subjects given consent for data collection?</li> </ul>

Analyses	<ul style="list-style-type: none"> <li>• How frequently are data transferred to the central database?</li> <li>• Can ad-hoc analyses be performed for the health care providers?</li> </ul>
Audit	<ul style="list-style-type: none"> <li>• How frequently are provider specific audit reports produced?</li> <li>• How frequently are multi-centre audit reports produced?</li> </ul>
Bibliography	<ul style="list-style-type: none"> <li>• List of related publications</li> </ul>
Management	<ul style="list-style-type: none"> <li>• Is the database approved by any clinical or professional bodies?</li> <li>• Who is involved in the management of the database?</li> <li>• Funding source</li> </ul>
Quality (rated from level 1 to 4)	<ul style="list-style-type: none"> <li>• Extent to which the population is representative</li> <li>• Completeness of the recruitment population</li> <li>• Variables included in the database</li> <li>• Completeness of the data (% complete for 95% of the data)</li> <li>• Form in which the data are collected (e.g., raw data)</li> <li>• Use of explicit definitions for variables</li> <li>• Use of explicit rules for deciding how data are recoded</li> <li>• Reliability of coding of conditions and interventions</li> <li>• Independence of observations to primary outcome</li> <li>• Extent to which data are validated</li> </ul>
Contact details	<ul style="list-style-type: none"> <li>• Name</li> <li>• Address</li> <li>• Telephone number</li> </ul>
Database classification	<ul style="list-style-type: none"> <li>• Body system</li> <li>• Pathogenesis</li> <li>• Intervention</li> <li>• Age group</li> <li>• Country</li> </ul>

Databases may be identified for inclusion in two ways. Database sponsors may submit their database to DoCDat for inclusion; alternately, DoCDat may become aware of a database and contact the owners to invite them to include it. DoCDat staff attempt to verify the information that is submitted about each database, but it is up to database sponsors to submit updates and keep their listing current.

Search capabilities are available; the database can be searched using type of patient, condition, treatment, and geographical area as criteria.

To be included in DoCDat, a database must meet the following criteria:

- Defines inclusion in the database as a common circumstance (e.g., an individual's condition, intervention required or undergone, administrative arrangement, or adverse outcome)
- Provides individual-level data, whether or not users of the database are permitted to know the identity of the individuals
- Includes data from more than one provider of health care

In addition, DoCDat includes some limited information on non-centralized individual level databases (meaning that the same data set is collected by more than one provider and not centrally collated) and aggregated databases that collect data on groups of people rather than individuals.

DoCDat does not store the actual data from each database; instead, it provides information on the database and contact information for the custodian. It is up to the interested researcher to contact the data custodian and attempt to make arrangements for use of the data.

Unlike Clinicaltrials.gov, inclusion of a database in DoCDat is purely voluntary. DoCDat currently has fewer than 200 databases registered.

### HSRProj

The Health Services Research Projects in Progress (HSRProj) database is a joint project of AcademyHealth and the Cecil G. Sheps Center for Health Services Research at the University of North Carolina at Chapel Hill. The project is funded by the National Library of Medicine. The purpose of the project is to provide access to ongoing research projects before the results are available in a published form. The main objectives of the database are to 1) provide up-to-date information on ongoing health services research projects; 2) identify individuals conducting cutting-edge health services research; 3) identify colleagues who may be interested in collaborating on a project or discussing implications of research; 4) identify funding sources, such as government agencies, private organizations, and foundations; and 5) identify specific ongoing health services research projects.<sup>17</sup>

The database includes over 7,500 health services research projects and is searchable. The projects include analyses of existing databases, analyses of linked data sources, and prospective observational studies, such as patient registries. The projects are funded by government agencies, foundations, and private organizations. This database does not meet the requirements for registration of controlled trials of FDAAA and ICMJE. The HSRProj database appears on the National Library of Medicine website.

The information included in HSRProj is similar to the information included in Clinicaltrials.gov, although not as detailed. The list of data elements included in the registry is provided in the table below:

**Table 4: HSRProj data elements**

Category	Data Elements
Investigator information	<ul style="list-style-type: none"> <li>Investigator name and contact information</li> <li>Location of study (institution and geographic location)</li> </ul>
Performing organization information	<ul style="list-style-type: none"> <li>Name of organization and contact information</li> <li>Supporting agency</li> </ul>
Funding	<ul style="list-style-type: none"> <li>Grant support</li> <li>Award type</li> <li>Initial and final year</li> </ul>
Abstract	<ul style="list-style-type: none"> <li>Text description of the study that generally includes clinical area of interest, study objectives, data sources, and methods</li> </ul>
MeSH terms	<ul style="list-style-type: none"> <li>Listing of relevant MeSH terms</li> </ul>
Project status	<ul style="list-style-type: none"> <li>Description of project status (e.g., ongoing, closed)</li> </ul>

Registration of health services research in the HSRProj database is not required by any regulations or legal means. Researchers can submit their information to the database using an online form. However, most of the information in the database was gathered by AcademyHealth. AcademyHealth prospectively contacts organizations to identify ongoing and new projects and then contacts the investigators to gather information for the HSRProj listing. The group developed a concept model of the key stakeholders in health services research and contacts these 150 stakeholders twice per year. Their efforts result in about 1,200 new project records being added each year. The project staff includes one full-time staff person, one full-time librarian, part-time oversight from the project head, and some administrative support.

**PDQ Cancer Clinical Trials Registry**

The Physician Data Query (PDQ) Clinical Trials Registry is a service of the National Cancer Institute (NCI) that includes over 8,000 open and over 19,000 closed cancer clinical trials and observational studies from within the U.S. and internationally. The site describes itself as the most comprehensive listing of clinical trials in cancer in the world. Trials and other studies for cancer treatment, genetics, diagnosis, supportive care, screening, and prevention are included. The trials and studies are sponsored by NCI as well as pharmaceutical companies, research hospitals, and other organizations. The registry meets the requirements for registration of controlled trials of FDAAA and ICMJE.

The information included in the PDQ Registry is similar to the information included in Clinicaltrials.gov. The list of data elements included in the registry is provided in the table below:

**Table 5: PDQ Cancer Trials Registry data elements**

Category	Data Elements
Basic trial information	<ul style="list-style-type: none"> <li>• Phase of the trial</li> <li>• Study type</li> <li>• Status (active, closed, etc.)</li> <li>• Age of patients</li> <li>• Sponsor</li> <li>• Protocol IDs</li> </ul>
Trial description	<ul style="list-style-type: none"> <li>• Summary of objectives</li> <li>• Description, including number of participating sites, number of enrolled patients, data collected, data collection intervals, and study duration</li> <li>• Disease, condition, drug, or therapy under study</li> <li>• Eligibility criteria</li> </ul>
Trial contact information	<ul style="list-style-type: none"> <li>• Lead organization/sponsor and contact information</li> <li>• Trial sites and contact information</li> </ul>
Links	<ul style="list-style-type: none"> <li>• Link to Clinicaltrials.gov listing</li> <li>• Statement of whether trial was registered initially with clinicaltrials.gov or with PDQ</li> </ul>

Clinicaltrials.gov and the PDQ registry are similar, but the databases are maintained separately. The databases are electronically linked such that information entered into PDQ is automatically entered into Clinicaltrials.gov in real-time and vice versa.

The PDQ registry appears on the NCI website and has NCI branding. The registry includes both health professional abstracts, which are written using medical terminology, and patient abstracts, which are written in non-technical language, whereas Clinicaltrials.gov generally only includes the health professional style abstract. The PDQ registry employs technical writers to develop the study descriptions and proactively contacts study investigators every three months to obtain updated information. Study investigators are required to review their study listing each year and make any necessary updates.

The registry has search capabilities that allow for searching by type of cancer, stage or subtype, type of trial, geographic location of the trial, type of treatment, drug name, trial phase, or a combination of these criteria. Online submission of trials is available through the NCI website.<sup>18</sup>

### **World Health Organization (WHO) International Clinical Trial Registry Platform (ICTRP)**

The ICTRP is a project of the WHO designed to facilitate clinical trial registration and reporting internationally. The project began following the Ministerial Summit on Health Research in 2004, during which participants called for the WHO to develop “a network of international clinical trials registers to ensure a single point of access and the unambiguous identification of trials.”<sup>19</sup>

The primary objective of the ICTRP is to facilitate the prospective registration of clinical trials and the public accessibility of trial information. Trials may be registered in Primary Registries and Partner Registries. Primary Registries meet WHO criteria for the content, quality and validity, accessibility, unique identification, technical capacity, and administration of the registry. Primary registries also meet the requirements of the ICMJE. Currently, Primary Registries are:

- Australian New Zealand Clinical Trials Registry (ANZCTR)
- Chinese Clinical Trial Register (ChiCTR)
- Clinical Trials Registry - India (CTRI)
- German Clinical Trials Register (DRKS)
- Iranian Registry of Clinical Trials (IRCT)
- ISRCTN.org
- Japan Primary Registries Network
- The Netherlands National Trial Register (NTR)
- Sri Lanka Clinical Trials Registry (SLCTR)

Partner Registries must meet the same WHO criteria as Primary Registries, but they do not need to be national in scope, supported by a government, managed by a non-profit agency, and open to all prospective registrants (i.e., they may be limited to a particular disease). Partner Registries do not have to meet the requirements for ICMJE, although some do. Currently, Partner Registries are the Physician Data Query (PDQ) Cancer Clinical Trials Registry, European Leukemia Trial Registry, Clinical Trial Registry

of the University Medical Center Freiburg, and German Registry for Somatic Gene-Transfer Trials (DeReG).

The WHO criteria for content include accepting all prospective registrations of interventional clinical trials submitted to the registry, being able to collect and publicly display data, endeavoring to keep the information up to date, and never removing a trial once it has been registered. The WHO criteria include having mechanisms in place to ensure the validity of submitted data, maintaining publicly accessible audit trails, and participating in the development of WHO Best Practice Guidelines for Clinical Trial Registries.

WHO also requires that trials registries allow the public to search their databases without having to pay any type of user fee or subscription fee, are electronically searchable, and are available for submissions and searching at any time on any day. The ICTRP requires that participating trials registries have processes in place to prevent the duplicate registration of trials and have mechanisms to facilitate the linkage of a trial that is registered in more than one registry. In addition to sufficient technical capacities and security provisions, WHO requires that its primary registries have support of a government within the country, publicly disclose ownership and governance structure, be managed by a non-profit, and agree to transfer all data to another WHO Primary Registry should the registry cease to function.

The data elements included in the WHO registries are listed in the table below:

**Table 6: WHO Primary/Partner Registry data elements**

Category	Data Elements
Summary data	<ul style="list-style-type: none"> <li>• Registry where study was originally submitted (e.g., Clinicaltrials.gov)</li> <li>• Date information last updated</li> <li>• Study ID</li> <li>• Date of initial registration</li> <li>• Primary sponsor</li> <li>• Study title</li> <li>• Date of first patient enrollment</li> <li>• Target sample size</li> <li>• Recruitment status</li> <li>• URL</li> <li>• Study type</li> <li>• Study design</li> </ul>
Countries of recruitment	<ul style="list-style-type: none"> <li>• List of countries where patients are recruited</li> </ul>
Contacts	<ul style="list-style-type: none"> <li>• Name</li> <li>• Address</li> <li>• Telephone number</li> <li>• Email</li> <li>• Affiliation</li> </ul>
Inclusion & exclusion criteria	<ul style="list-style-type: none"> <li>• List of inclusion/exclusion criteria</li> </ul>
Health condition or problem	<ul style="list-style-type: none"> <li>• Description of health condition or problem under study</li> </ul>

studied	
Interventions	<ul style="list-style-type: none"> <li>• Description of study interventions</li> </ul>
Outcomes	<ul style="list-style-type: none"> <li>• Description of primary and secondary outcomes</li> </ul>
Funding sources	<ul style="list-style-type: none"> <li>• List of primary/secondary funding sources</li> </ul>

The WHO site includes a search portal where it is possible to search for trials. The site is considering whether to include study results.

### Comparison of Existing Databases

Table 7 below compares the key information for each of the five existing databases.

**Table 7: Comparison of Existing Databases**

	Clinicaltrials.gov	DoCDat	WHO	PDQ Cancer Trials Registry	HSRProj
How are registries identified for inclusion in the database?	Registry sponsors self-identify and submit registries to Clinicaltrials.gov	Registry sponsors may self-identify and submit their information, or they may be contacted by DoCDat and invited to participate	Registry sponsors self-identify and submit registries to a partner or primary registry	Registry sponsors self-identify and submit registries to the database	Staff prospectively contact organizations and identify projects for inclusion
Voluntary or required?	Mostly voluntary for registries, but can be used to fulfill U.S. legal requirements for some trials	Voluntary	Voluntary	Voluntary	Voluntary
What incentives do patient registries have to submit information?	Concerns regarding ability to publish results	Interest in facilitating research collaboration	Concerns regarding ability to publish results	Concerns regarding ability to publish results	Interest in facilitating research collaboration

	Clinicaltrials.gov	DoCDat	WHO	PDQ Cancer Trials Registry	HSRProj
How is the information verified?	All clinical study information submitted to ClinicalTrials.gov is reviewed by the National Library of Medicine	DoCDat staff attempt to verify all submitted information before posting	Requires that participating databases have mechanisms in place to ensure the validity of submitted data and maintain publicly accessible audit trails	Information is reviewed by National Cancer Institute staff; descriptions of trials are written by staff and reviewed annually by investigators	Information is reviewed by AcademyHealth and National Library of Medicine staff
How are the databases kept current?	It is the responsibility of the investigators who submitted the studies to keep the listings current	Database owners can submit updates to DoCDat, as needed	Requires that participating databases endeavor to keep their information up to date, and never remove a trial once it has been registered	Investigators must review study listing annually; investigators receive email reminders to update study information every three months; updates can be submitted at any time	Database staff contact sponsors every six months to update information
How are the databases indexed for searching?	By study location and topic, but advanced search options are also available (study type, sponsor, subject population, etc.)	By type of patient, condition, treatment, and geographical area	Search criteria vary by database (primary and partner registries each have their own criteria)	By type of cancer, stage or subtype, trial type, location, type of treatment, drug name, and trial phase	Searching is available by key words, by investigator, by funding organization, or by state

**Relevance of Existing Databases for a Registry of Patient Registries**

While none of the existing databases is designed to achieve exactly the same objectives envisioned in this paper for a registry of patient registries, these databases provide both relevant models for discussion as well as methods to address specific requirements. Clinicaltrials.gov provides an excellent model for the study-listing component described in Table 1. Even as currently listed, standard data

elements are very close to what might be envisioned for a registry of patient registries. Ideally, such a list would be further expanded to include some of the additional requested components, specifically a voluntary listing of data elements, code lists, and so forth. The search functionality of Clinicaltrials.gov is straightforward, and a similar functionality would work well for a registry of patient registries. The results reporting framework provides an infrastructure that could be similarly used for reporting registry results. While the table structures would differ, overall storage and search functions would likely be similar. In terms of serving as a potential data repository, Clinicaltrials.gov is not currently designed to serve this role and does not have such functionality.

Adapting Clinicaltrials.gov for several or potentially all of the functions for a registry of patient registries is conceivable. As noted above, the listing and search components are already sufficient and modifications to the framework, such as listing data element types, could be made in a new patient registry module. Posting via this vehicle would also eliminate the redundancy of potentially needing to post in more than one place (e.g., Clinicaltrials.gov and a new registry of patient registries). The results reporting framework potentially could be adapted to the types of results useful for patient registries. Table structures might differ, but the search capabilities would likely be the same. The strengths and limitations of this approach are discussed in more detail below, under the “Collaborate with Clinicaltrials.gov” section.

DoCDat similarly offers some useful insights into developing a registry of patient registries. While it is far less sophisticated in search and listing functionality to Clinicaltrials.gov, the DoCDat system provides a good outline of the kind of information that might be useful to a researcher looking for existing registries as sources of data or for collaboration. While the list of data assets and self-reported quality information are far more extensive than most registry developers would likely be willing to complete and update on a regular basis, it provides a very useful starting point for designing such a framework for AHRQ’s purposes.

The HSRProj database is not strongly applicable to designing a registry of patient registries. It does not provide the depth of information that would be included in a registry of patient registries, nor does it offer the search capabilities that would be needed for this project. However, the HSRProj database’s facilitated model for collecting data using active outreach may be applicable to the registry of patient registries. This model is discussed in more detail in the “Incentives” section below.

The PDQ Cancer Trials Registry accomplishes several goals that are relevant to a registry of patient registries. While it is focused on cancer trials, the system leverages an XML interchange of data with Clinicaltrials.gov that eliminates potential duplication of effort by the developers listing the trial and still allows PDQ Cancer Trials Registry to maximize its own objectives through its website. These objectives include a primary goal of increasing participation in cancer trials that is assisted by the inclusion of more patient relevant information, such as patient-appropriate descriptions of the trial, on their website. The approach utilized by the PDQ system is a hybrid model, which leverages Clinicaltrials.gov while maintaining certain independent functionalities. This approach is explored in more detail below, under the “Hybrid” section.

The WHO system is not truly applicable to designing a registry of patient registries, but it does paint a picture of the complexities of potentially expanding a registry of patient registries internationally. It is clear that this would be a major undertaking.

## **PROPOSALS FOR REGISTRY OF PATIENT REGISTRIES**

Using the information gained from the interviews with stakeholders and the review of existing databases, the following questions can now be addressed:

1. Should there be a registry of patient registries?
2. What are the requirements for the registry of patient registries?
3. What approaches might AHRQ pursue to develop and maintain a registry of patient registries?
4. What would motivate researchers to list their patient registries in the registry of patient registries (i.e., incentives)?
5. What are the costs of setting up and maintaining a registry of patient registries?

### **Should There Be a Registry of Patient Registries?**

The clear finding from stakeholder interviews is that there is broad interest in a registry of patient registries that would serve several functions that are not sufficiently served by existing programs, such as Clinicaltrials.gov. As described earlier, there would be incremental gain in each additional function supported by such a registry. First, the searchable listing of existing registries in sufficient detail with contact information would help avoid duplication of efforts, provide opportunities for collaboration, increase transparency, and potentially improve participation by providers and patients. Adding a data element listing would improve standardization and potentially enable linkage or aggregations to be performed more easily in the future. A results database would not only link to publications but could provide potentially useful information on populations and diseases that might otherwise not be accessible. A detailed description of data assets would be a boon to other researchers seeking data sets to answer new questions or for linkage studies. While a repository of data would be more complicated to administer, and there are no current models to leverage, it could provide a place to store data (with appropriate controls) for use after a registry ends.

Equally clear, however, is the fact that without a strong incentive, such as a mandate from ICMJE or a legal requirement, voluntary postings of registries in the database are likely to be limited. The experiences of DoCDat in having registries voluntarily submit their data confirms this concern. HSRProj has had some success in overcoming this barrier by reaching out to investigators and actively identifying studies for inclusion (albeit at additional cost). While there is likely to be value in even an incomplete listing of patient registries, such value would need to be weighed against the costs of different approaches for developing and maintaining the registry of patient registries.

### **What Are the Requirements for the Registry of Patient Registries?**

The requirements for the registry of patient registries were developed based on interviews with stakeholders and research into the existing databases. These requirements are described in detail in Table 1. To summarize, the registry of patient registries is envisioned as a web-accessible database

containing standard information on each patient registry. The database would be available at all times to any interested party (the public, government employees, researchers, etc.). The website would have a mechanism to allow registry sponsors to set up accounts and add new registry listings or update existing registry listings.

The registry of patient registries could serve up to four main objectives. First, its primary purpose would be as a list of existing patient registries. For each included registry, information on the registry purpose, study design, disease area, patient population, participating sites, enrollment target, status of enrollment, geographic location, and data collection would be available in a standard format. Second, the site would enable visitors to search for registries using a range of data elements (e.g., study design, disease area, geographic location, etc.). The search engine would also use the MeSH hierarchy of terms to help guide users to appropriate studies. Third, summaries of registry analyses could be stored and searchable in standard table formats. Links to PubMed for any publications resulting from the registry could also be presented. Fourth, the site could include a description of collected data assets with information to facilitate collaboration. This would include both active and closed patient registries. In addition, for registries that close and choose not to maintain their data assets, a repository of data could potentially be built and maintained by AHRQ. Information on these assets would also be accessible via the website. As previously discussed, these objectives have different levels of priority among stakeholders and would also have different levels of compliance by registry owners. As a result, the listing and search functions are the key functionality, while the remaining two functions contribute to opportunities for collaboration and knowledge sharing.

Validation of the listing information will be an important component of a registry of patient registries. Several stakeholders expressed concerns about ensuring that the information provided in the database is accurate and complete. The registry of patient registries will need to have a transparent process in place to review the information that is submitted to the database for accuracy and to ensure that updates are made to information as necessary (e.g., on a yearly basis). Staff of the registry of patient registries could perform the initial validation by contacting the principal investigator or reviewing published literature to verify key facts. In terms of keeping information up to date, registry sponsors could be required to verify that their information is correct on a yearly basis in order to maintain the registry listing.

### **What Approaches Might AHRQ Pursue to Develop and Maintain a Registry of Patient Registries?**

Using these requirements as a guide, three approaches are proposed for how AHRQ might develop and maintain a registry of patient registries. Interviews were conducted with representatives of Clinicaltrials.gov, DoCDat, HSRProj, and AHRQ to inform the approaches described below.

#### **Proposal 1: Develop the Registry of Patient Registries *De Novo***

Using the requirements described in this paper, AHRQ could develop a registry of patient registries *de novo* either internally using its existing infrastructure (National Resource Center for HIT) or as an outsourced contract to an independent center. The discussion below focuses on the former alternative as it is likely to be less expensive given the significant investments already made in the AHRQ

infrastructure. AHRQ currently has database capabilities, website capability, and collaborative tools. In addition to the listing component, AHRQ has secure portal capabilities that would allow users to securely upload and update data and to upload files. In other words, registry data sets could be stored. Primarily, AHRQ would utilize its contracting capability and existing infrastructure to build a system that would emulate many of the features of Clinicaltrials.gov in terms of listing and search requirements and design and test new requirements for the differentiating features of this project. With this approach, AHRQ could use its branding on the site and select a site name that clearly relates to the purpose of the database (e.g., registries.ahrq.gov, patientregistries.gov, etc.).

While this is certainly achievable, there are several limitations to this approach. First, because this would be an independent program separate from Clinicaltrials.gov, any registry that chose or was required to submit to Clinicaltrials.gov would need to duplicate effort to maintain information in both places. Second, the search and listing components would need to be developed while they already exist on Clinicaltrials.gov; this could be viewed as an unnecessary use of resources within Health and Human Services (HHS). Third, searchers would need to search in two places (the new website and Clinicaltrials.gov) to identify registries that could be listed through either program. Fourth, registries choosing to list on the AHRQ database would not benefit from the relatively broad exposure (e.g., to patients) gained from participation on Clinicaltrials.gov. Fifth, journal reviewers would need to be educated to also look at the AHRQ database when reviewing manuscripts from registry developers (note, even though registries are not generally required to list, journal reviewers commonly search on Clinicaltrials.gov to gain background information).

### **Proposal 2: Collaborate with Clinicaltrials.gov to Build an Integrated Patient Registry Module**

A second approach is to provide support to Clinicaltrials.gov to meet the requirements of the registry of patient registries. The Clinicaltrials.gov website is both user friendly and has a proven record of accomplishment in providing accurate listings and cross references through the NLM system. As noted earlier, the listing requirements for the registry of patient registries would be very similar to those currently used by Clinicaltrials.gov. Clinicaltrials.gov would essentially create a registry module that would be integrated into the existing searching system. The patient registry module would need to include some additional registry-specific information as well as a potential listing of data elements. Searching for registries could be facilitated with the creation of a specific category for patient registries. Although still incomplete, the results database could be customized for registries and leverage the same Clinicaltrials.gov engine. In addition to leveraging existing and effective infrastructure, use of Clinicaltrials.gov would eliminate the duplicate-listing problems described under Proposal 1.

However, there are several limitations to this approach as well. First, this would be an additional function for Clinicaltrials.gov that may be lower on its priority list than other functions required, for example, under FDAAA. Second, it may be confusing to users to have different functions available depending on whether patient registry developers listed their patient registry as a registry versus a different designation, such as a prospective observational study. It is unclear what effort would be required for AHRQ and Clinicaltrials.gov to develop a common lexicon of terms that would refer users back to the patient registry module. Third, Clinicaltrials.gov currently has neither a listing of data assets

nor a repository of data. Should these functionalities be included by AHRQ, they would need to be built. In particular, Clinicaltrials.gov does not have the infrastructure required to allow raw data to be stored securely and confidentially for the purposes described above and likely would not be interested in adding this capacity.

### **Proposal 3: Develop Hybrid Approach that Leverages Capabilities of Clinicaltrials.gov & AHRQ**

The third approach follows a variation of the model used by PDQ Cancer Trials Registry. In this approach, AHRQ would leverage Clinicaltrials.gov for specific capabilities (e.g., search and listing) and avoid the problem of duplicate databases through bidirectional feeds. At the same time, AHRQ would develop and maintain its own infrastructure to support requirements not met by Clinicaltrials.gov such as the data assets listings and potential data repositories with different levels of access. Another advantage of this approach is that AHRQ could develop this in stages and potentially make decisions on subsequent investment based on the level of registration by registries. As noted earlier, incentives will be very important to successfully registering enough patient registries in different conditions to demonstrate value. The first stage would focus on the listing and search component, which would require the least investment. If there were reasonable participation by registries, AHRQ could plan, develop, and implement the next phases over time. AHRQ could also exert more control over the branding and name of the registry of patient registries.

The limitations of this approach are primarily in the coordination it would require between AHRQ and Clinicaltrials.gov. A potential risk is the level of priority (relative to the requirements under FDAAA that Clinicaltrials.gov must meet) that would be accorded to these efforts.

### **What Would Motivate Researchers to List Their Registries in the Registry of Patient Registries?**

In addition to choosing an approach, AHRQ must consider how to incentivize registry sponsors to submit their information to the registry of patient registries. As discussed above, registration of patient registries is not mandated currently by law or publishing requirements. Answering the question of incentives is critical to developing a functional, effective registry of patient registries. Potential participation 'incentives' could range from purely voluntary to truly mandatory.

From a voluntary perspective, registry owners might be motivated to participate based on goals such as: to be transparent; to seek collaborators for their registry; to aid in recruitment of patients and providers to their registries; to market the availability of their data assets for use or collaboration; or to ensure the continued use of the data for the public good even when the registry otherwise ends. With any of the proposals mentioned above, registries could be identified and included in two ways: 1) registry sponsors submit their information to the database, or 2) the database contacts organizations to identify and include registries. Under the first approach, registry sponsors would be the party responsible for deciding to list their registry, completing and submitting information on their registry to the database, and updating that information as needed. The process would need to be as straightforward as possible to minimize the perceived burden of participating.

Voluntary participation alone, however, may not be effective, as seen with DoCDAT. Additional incentives could take many forms ranging from incremental benefits of listing to mandates from editors or potentially legal or regulatory requirements. With respect to incremental benefits, the website could offer additional capabilities only to registry owners who post—such as data element or results searching. Stronger incentives might include making posting a requirement of receiving HHS funding, working with journal editors towards a listing requirement, or creating a regulatory requirement for registries sponsored by the regulated industries. Since a large number of registries are not in the regulated sector and many manufacturers already post the registries that they sponsor on Clinicaltrials.gov, the latter would have less of an impact. Similarly, tying HHS grant funding to a posting would have a strong impact, but only for those registries that have grant funding. Since publications are the most common driver for patient registries, tying participation to the likelihood of journal publication would be helpful in incentivizing participation. Even if the ICMJE could not be convinced to adopt such a position, obtaining the buy-in of a few of the more prestigious journals could have a significant impact on interest in registering.

The lack of a perceived mandate might be ameliorated to some degree if the registry of patient registries proactively contacts registry sponsors and collects information to complete registry listings in the database. This approach is similar to that used by the HSRProj database. In that model, the database team developed a list of key stakeholders in health services research. The team contacts the stakeholders twice per year to learn of new research projects and then contacts the principal investigator for the research to obtain the necessary information to list the project in the database. AHRQ could take a similar approach by implementing a direct outreach program aimed at medical societies, patient advocacy organizations, academic centers, and Federal research centers. This outreach program could be used to identify registry sponsors. Automated tools might be used to notify the registry owner at least yearly to update the listing. This approach would significantly reduce the burden of participation on registry sponsors, which was the key barrier cited in interviews with stakeholders.

However, this approach is costly. The HSRProj effort utilizes a full time employee for this function, and this is potentially a larger effort. It is possible, however, that AHRQ outreach activities may become less important as the database grows in size and scope. Researchers may recognize the value of the database and voluntarily submit their information or update their information, without prompting from AHRQ. AHRQ could also consider moving ahead with the outreach approach as an interim plan while seeking other incentives discussed above.

### **What Are the Costs of Setting Up and Maintaining a Registry of Patient Registries?**

Without a full evaluation of each requirement with respect to cost paid by AHRQ or Clinicaltrials.gov for similar work, cost information can only be generally estimated. However, it is clear that the options listed do have different cost implications. Proposal 1, building a *de novo* registry of patient registries, is the most expensive option. Based on discussions with Clinicaltrials.gov, AHRQ would likely need several

million dollars in start-up funding to replicate the search and listing components and then more than \$1 million per year to maintain the system.

Proposal 2, leveraging the Clinicaltrials.gov infrastructure to create a patient registry module, would be the least expensive approach. Clinicaltrials.gov believes around \$1 million would be required to build the unique interface requirements, and approximately one to two additional staff members would be needed to maintain the registry of patient registries on the Clinicaltrials.gov website. For the results database component, the new Clinicaltrials.gov database could also be leveraged and customized to provide unique tables for registries. The current Clinicaltrials.gov results database has a budget of about \$5 million per year, including a staff of approximately 20 persons.

Proposal 3, using a hybrid approach, will have an intermediate cost depending on which components are developed and maintained by AHRQ. An advantage of both this approach and the full Clinicaltrials.gov approach are that the listing and search components can be accomplished rapidly, without fully investing in new infrastructure, and the impact of these components can be evaluated in making decisions about further investments (e.g., a results database, data repository, etc.).

In addition to the direct funding by AHRQ, there are other potential financing options. These include seeking financial support of additional public and private entities that might benefit from the registry of patient registries (e.g., CMS, FDA, medical professional societies, manufacturers). External support will be easier to obtain once success has been demonstrated in registering patient registries.

Based on discussions with existing trials registries and AHRQ, rough estimates of the relative costs of setting up the listing and search components using each approach were developed. This information should be used cautiously since requirements are not fully detailed for the website, and the data used was not verified using a competitive bidding model. Also, information is not available to estimate the costs of the results database or an asset listing component. Information on the cost of pursuing the various incentives described above is also not included here. This available cost information is summarized in Table 8 below.

**Table 8: Estimated Costs by Proposal of Set-Up and Maintenance for Registry Listings and Search Capabilities**

	<i>De Novo Registry</i>	<i>Clinicaltrials.gov “Add-on” Model</i>	<i>Hybrid Model</i>
<b>Set-Up</b>	\$\$\$	\$\$	\$\$
<b>Annual Maintenance</b>	\$\$	\$	\$

*\$ = less than \$500,000; \$\$ = \$500,000 to \$2,000,000; \$\$\$ = greater than \$2,000,000*

## CONCLUSIONS

A review of the existing databases for trials registration and interviews with stakeholders clearly demonstrated that there is an unmet need for a registry of patient registries. Government agencies, medical associations, foundations, and health services researchers all noted that potential value of a

registry of patient registries and identified several key objectives of such a system. Databases such as Clinicaltrials.gov, DoCDat, HSRProj, and the PDQ Clinical Trials Registry do not currently meet these objectives.

Despite the broad interest in a registry of patient registries, a key issue is the lack of current incentives for registry owners to submit their information to such a database and update it in a timely fashion. Developing incentives or pressures to list patient registries will be an important success factor for such a project. Clinicaltrials.gov did not see significant participation until U.S. law and ICMJE policies mandated participation. It is likely that an AHRQ-sponsored registry of patient registries would also have difficulty encouraging widespread voluntary participation without external pressure on registry owners from journal editors, agencies that fund registries, or agencies that regulate registries.

While none of the existing databases reviewed here exactly meets the full list of requirements gathered from the stakeholders, they provide helpful information on how AHRQ might design and support the implementation of such a registry of patient registries. Given the participation risks described above, the most compelling approach may be one that begins with the core functionality of listing and search and builds other components over time. The least expensive strategies are those that leverage existing systems such as Clinicaltrials.gov. Enabling bidirectional flow of information from a registry of patient registries to Clinicaltrials.gov would be important to minimize duplication of effort by investigators who seek to have information reside in both places. A hybrid model that uses Clinicaltrials.gov for listing and search functions and independently hosts components for other functions is one means of accomplishing both requirements, but there are other viable models as presented above.

Further interviews and discussions with both stakeholders and registry owners are highly recommended to help AHRQ further define requirements, priorities, incentive structures, and funding options in planning next steps.

## REFERENCES

---

- <sup>1</sup> Zarin DA, Ide NC, Tse T, et al. Issues in the Registration of Clinical Trials. *JAMA*. 2007;297(19):2112-2120.
- <sup>2</sup> Food and Drug Administration Modernization Act of 1997, Pub L No. 105-115, 1997.
- <sup>3</sup> "About Clinicaltrials.gov." Clinicaltrials.gov. National Institutes of Health. U.S. Department of Health and Human Services. Accessed online on May 11, 2009 at: <http://Clinicaltrials.gov/ct2/info/about>.
- <sup>4</sup> Food and Drug Administration Amendments Act of 2007, Public Law 110-85. Signed into law in September 2007. See Title IX, Section 905.
- <sup>5</sup> DeAngelis CD, Drazen JM, Frizelle FA, et al. Is this clinical trial fully registered?—a statement from the International Committee of Medical Journal Editors. *JAMA*. 2005;293:2927-2929.
- <sup>6</sup> Zarin DA, Ide NC, Tse T, et al. Issues in the Registration of Clinical Trials. *JAMA*. 2007;297(19):2112-2120.
- <sup>7</sup> "Clinical Trial Registration: Looking Back and Moving Ahead." June 2007 Update on Trials Registration. International Committee of Medical Journal Editors. Accessed online on May 11, 2009 at: [http://www.icmje.org/clin\\_trial07.pdf](http://www.icmje.org/clin_trial07.pdf).
- <sup>8</sup> Zarin DA, Tse T. "Moving Toward Transparency of Clinical Trials." *Science*. 2008 March 7; 319(5868): 1340–1342.
- <sup>9</sup> Moher D. Clinical-trial registration: a call for its implementation in Canada. *CMAJ* 1993;149[11]:1657–8.
- <sup>10</sup> Gliklich RE, Dreyer NA, eds. Registries for Evaluating Patient Outcomes: A User's Guide. (Prepared by Outcome DEClDE Center [Outcome Sciences, Inc. dba Outcome] under Contract No. HHS290200500351 TO1.) AHRQ Publication No. 07-EHC001-1. Rockville, MD: Agency for Healthcare Research and Quality. April 2007.
- <sup>11</sup> Google. Accessed online at: <http://www.google.com/>.
- <sup>12</sup> Google Scholar. Accessed online at: <http://scholar.google.com/>.
- <sup>13</sup> PubMed. Accessed online at: <http://www.ncbi.nlm.nih.gov/pubmed/>.
- <sup>14</sup> Guidance for Industry: Information Program on Clinical Trials for Serious or Life-Threatening Diseases and Conditions. Food and Drug Administration. U.S. Department of Health and Human Services. March 2002. Accessed online on May 14, 2009 at: <http://www.fda.gov/cder/guidance/4856fnl.htm>.
- <sup>15</sup> Clinicaltrials.gov Fact Sheet. National Library of Medicine. U.S. National Institutes of Health. Accessed online on May 14, 2009 at: <http://www.nlm.nih.gov/pubs/factsheets/clintrial.html>.
- <sup>16</sup> N. Black and M. Payne on behalf of the DoCDat Development Group. Directory of clinical databases: improving and promoting their use. *Qual Saf Health Care* 2003; 12:348-352. Accessed online on November 11, 2008 at [www.qshc.bmj.com](http://www.qshc.bmj.com).
- <sup>17</sup> HSRPROJ Fact Sheet. National Library of Medicine. U.S. National Institutes of Health. Accessed online on May 28, 2009 at: <http://www.nlm.nih.gov/pubs/factsheets/hsrproj.html>.
- <sup>18</sup> PDQ Cancer Clinical Trials Registry. PDQ® - NCI's Comprehensive Cancer Database. National Cancer Institute. U.S. National Institutes of Health. Accessed online on May 14, 2009 at: [http://www.cancer.gov/cancertopics/pdq/cancerdatabase#clinical\\_trial](http://www.cancer.gov/cancertopics/pdq/cancerdatabase#clinical_trial).

---

<sup>19</sup> About the WHO ICTRP. International Clinical Trials Registry Platform (ICTRP). World Health Organization. Accessed online on May 14, 2009 at: <http://www.who.int/ictcp/about/en/>.