



**THE LISTER HILL NATIONAL CENTER  
FOR BIOMEDICAL COMMUNICATIONS**

*A research division of the U.S. National Library of Medicine*

---

**TECHNICAL REPORT  
LHNCBC-TR-2005-003**

**ClinicalTrials.gov;**  
A Report to the Board of Scientific Counselors

May 2005

Investigator: Deborah A. Zarin, M.D., Office of the Director  
Annice Bergeris  
Nicholas C. Ide, M.S.  
Tony Tse, Ph.D.

---

U.S. National Library of Medicine, LHNCBC  
8600 Rockville Pike, Building 38A  
Bethesda, MD 20894



## **Executive Summary**

ClinicalTrials.gov was established by Section 113 of the Food and Drug Administration Modernization Act of 1997 (FDAMA 113) and launched by the National Institutes of Health in February 2000. The site currently includes information on nearly 13,000 studies for hundreds of diseases and conditions conducted in about 100 countries. Sponsors representing the Federal government, pharmaceutical industry, and non-profit organizations from around the world submit and maintain study information using the National Library of Medicine-developed Protocol Registration System.

Since the passage of FDAMA 113, and particularly in the past year, many groups, including Congress, have begun to realize that clinical trial registries may fulfill other critical needs. In particular, growing awareness that some trial results are never published, or are published only in part, have led to concerns that patients and policymakers who rely on a complete understanding of all of the research on a particular topic may be consistently misled. Various groups have called for the expansion of ClinicalTrials.gov, including the American Medical Association, the International Committee of Medical Journal Editors (ICMJE), the World Health Organization, and representatives of the pharmaceutical industry (e.g., PhRMA). As a result of the public attention and debate, ClinicalTrials.gov has expanded its scope and is in ongoing discussions with key stakeholders regarding possible other changes.

To place these needs in perspective, we present a conceptual framework of clinical trial registries that describes five primary purposes of registries cited in the literature and the data elements required for each. Next, the scope of the registries is discussed in the context of types of clinical research studies and sponsors, as called for by various policy initiatives. Incorporation of the ICMJE criteria illustrates the most recent expansion of ClinicalTrials.gov to accommodate a rapidly changing environment. The challenges of assuring data quality on information provided to ClinicalTrials.gov by external sources are discussed. The implementation of the ClinicalTrials.gov system is then described along with its context within the Lister Hill National Center for Biomedical Communications. Future challenges include developing mechanisms to improve registration rates, ensuring the accuracy of information submitted to ClinicalTrials.gov, and evaluating the overall effectiveness of the site.

ClinicalTrials.gov  
A Report to the Board of Scientific Counselors  
May 5, 2005

**Index**

I. Introduction and Background .....	1
A. Impetus	
B. Data Content	
II. Scientific Issues.....	4
A. Conceptual Framework	
B. Data Elements	
C. Types of Trials	
III. Policy Landscape .....	8
A. Federal Legislation	
B. International Committee of Medical Journal Editors (ICMJE)	
C. World Health Organization (WHO)	
D. Pharmaceutical Industry	
E. State Legislation	
F. Cochrane Collaboration	
IV. Data Quality Issues .....	12
A. The Denominator Problem: Definition and Compliance	
B. The Verification Problem	
C. Current Quality Assurance Procedures	
V. Implementation .....	14
A. System Design	
B. Research and Development	
VI. Future Directions/Challenges .....	19
A. Improve Registration Rates of Relevant Trials	
B. Evaluate the Utility and Feasibility of Reference Protocols	
C. Pilot Study of Clinical Trial Verification Procedures	
VII. References .....	21
Appendix A. Data Elements for ClinicalTrials.gov .....	23
Appendix B. Questions for the Board.....	28

**I. Introduction and Background**

ClinicalTrials.gov<sup>1</sup> was established by Section 113 of the Food and Drug Administration (FDA) Modernization Act of 1997 (FDAMA 113, 1997) and launched by National Institutes of Health (NIH) in February 2000. The site was designed to provide the public with information about ongoing and completed clinical trials. ClinicalTrials.gov currently includes information on approximately 13,000 studies for hundreds of diseases and conditions conducted in about 100 countries. Sponsors representing the Federal government, pharmaceutical industry, and non-profit organizations from around the world submit and maintain study information using the National Library of Medicine (NLM)-developed Protocol Registration System (PRS)<sup>2</sup> (Gillen et al., 2004).

---

<sup>1</sup> <http://clinicaltrials.gov/>

<sup>2</sup> <http://prsinfo.clinicaltrials.gov/>

ClinicalTrials.gov uses many of the medical informatics tools and services developed at NLM. In addition, the project contributes to research and development products at the Lister Hill National Center for Biomedical Communications (LHNCBC). For example, automatically creating links from studies described in ClinicalTrials.gov to relevant health topics at MedlinePlus<sup>3</sup>, is possible through the NLM Medical Subjects Heading<sup>®</sup> (MeSH<sup>®</sup>) controlled vocabulary thesaurus. Automatically expanding a patient's search request with synonyms, such as "myocardial infarction" for "heart attack," relies on the Unified Medical Language System<sup>®</sup> (UMLS<sup>®</sup>). Ongoing research and development by ClinicalTrials.gov has resulted in new technologies and products used by other NLM projects.

From the outset, ClinicalTrials.gov has been designed for use by patients. Many technologies were incorporated to aid patients in finding and understanding clinical research studies, such as help with technical terminology and resource information on understanding clinical trials. NLM continues to explore improved access methods, evaluate different ways people use consumer health information, and to classify types of questions asked by users. Continuing research into areas such as consumer health information needs (e.g., Tse and Logan, submitted) and online seeking behaviors (McCray et al., 2004), methods for assessing reading ease in the health domain (Gemoets et al., 2004), and Spanish-English bilingual cross-language information retrieval (Roseblat et al., 2004; submitted), will likely have a direct impact on improving access to ClinicalTrials.gov data for all.

### **A. Impetus**

The FDA Modernization Act, Section 113 (FDAMA 113) requires that clinical trials for effectiveness conducted under an investigational new drug application (IND) for "serious or life threatening diseases" must be registered in ClinicalTrials.gov. Specifically, the law focuses on the goal of helping potential subjects find clinical trials in which to participate. Since the passage of that law, and particularly in the past year, many groups, including Congress, have begun to realize that clinical trial registries may fulfill other critical needs. In particular, growing awareness of the fact that some trial results are never published, or are published only in part, have led to concerns that patients and policymakers who rely on a complete understanding of all of the research on a particular topic may be consistently misled. This has led to calls for the expansion of ClinicalTrials.gov. As a result of the public attention and debate, ClinicalTrials.gov has expanded its scope and is in ongoing discussions with key stakeholders regarding possible other changes. These issues will be discussed further in Section III, Policy Landscape.

### **B. Data Content**

ClinicalTrials.gov includes trials that are mandated by FDAMA 113, as well as other trials that are voluntarily submitted. At this time, the registry includes a broad range of studies and is not constrained by any one definition of *clinical trial*. Current summary statistics illustrating the range of data content, intervention types, and sponsors are shown below (Tables 1, 2, 3).

---

<sup>3</sup> <http://medlineplus.gov/>

Overall Status	Total	Study Type	
		Interventional	Observational
Not yet recruiting	244	213	31
Recruiting	4,322	3,590	732
No longer recruiting	4,034	3,640	394
Suspended	209	205	4
Terminated	181	170	11
Completed	3,936	3,001	935
Total	12,926	10,819	2,107

Table 1. Overall Recruitment Status of All Records in ClinicalTrials.gov (4/14/05)

Intervention Type	Total
Drug	9,266
Procedure	3,896
Behavioral	627
Vaccine	276
Device	151
<i>Not Provided</i>	2,180
Total	16,396

Table 2. Intervention Types Represented in ClinicalTrials.gov (4/18/05).  
NOTE: Total exceeds number of trials because trials have multiple arms.

Lead Sponsor Type	No. of Records	No. of Organizations
NIH (institutes and centers)	6,492	21
University/Other	3,245	345
Pharmaceutical Companies	2,746	463
Federal, non-NIH	411	6
International	46	34
Individual	11	10
Total	12,950	880

Table 3. Number of Sponsors Represented in ClinicalTrials.gov (4/15/05)

Initially, ClinicalTrials.gov only accepted information about studies conducted under FDA's IND regulations or funded by the U.S. Federal government. Restricting data sources in this way allowed ClinicalTrials.gov to rely on Federal partners to corroborate (at least in theory) information provided by study sponsors. Since October 2004, ClinicalTrials.gov expanded its policies to allow the registration of any study that meets the following requirements: (1) approval by a human subject review board (or equivalent) and (2) conformance with the regulations of the appropriate national or international health authority. A mechanism for corroborating this information is a significant challenge and several approaches are being investigated.

In addition to entries for specific trials, the public site includes background materials about clinical trials (e.g., “An Introduction to Clinical Trials”) as well as links to relevant resources. Users may search and browse by location, condition, and/or other trial attributes (Tse, Johnson, & Ripple, 2002).

## II. Scientific issues

### A. Conceptual Framework

Clinical trial registries serve a number of purposes for a variety of stakeholders. Table 4 illustrates the different functions of trial registries as reported in the literature, along with the type of information required to fulfill those functions.

		Data Element Types				
		Administrative	Protocol	Event Log	Recruitment	Results
Registry Type/Purpose	Existence	●	●			
	Patient Matching	●	●		●	
	Verification	●	●	●		
	Systematic Review	●	●	●		●
	Public Access	●	●			●

Table 4. Clinical Trial Registries Framework

**1. Existence.** At a fundamental level, a registry serves to record the existence of a trial and some basic information about the trial. To provide information on the existence of a trial, assigning a persistent, unique trial identifier to each trial is critical. "...This requirement...can be compared with use of the International Standard Book Number" (ISBN) (Dickersin & Rennie, 2003). We refer to such requirements collectively as "administrative" data elements.

Existence registries allow users to determine what trials exist (uniquely), based on some specific attribute(s) (e.g., condition being studied), and provide users with a pointer to additional information. Using a comprehensive existence registry, stakeholders such as funders or researchers could quickly determine the numbers of trials conducted in a particular area, but would need to spend considerable time finding other relevant information about the study, such as status and, if completed, the results.

**2. Patient Matching.** Another function of a trial registry is to match eligible participants with relevant clinical trials. This function helps patients, their care givers, and trial investigators. In order to serve this function, a registry needs additional information about the trial protocol (e.g., condition under study, intervention being evaluated, etc.) as well as up to date information about trial sites that are actively recruiting subjects. These recruitment data elements are a particular challenge for data providers as they need to be maintained and updated throughout the life cycle of the trial.

**3. Verification.** Verification refers to the need for those who evaluate clinical trials to determine whether the reports of a trial accurately reflect the initial intent and the conduct of the trial. For example, a journal editor may want to verify that the outcome measures reported in a manuscript are the outcome measures that were identified a priori (this is important both to ensure that the statistical measures are appropriate to the initial trial design, and to ensure that only "positive" outcomes aren't selectively culled for reporting from all of the outcome measures). Key dates may also be important for this purpose. For example, evaluators may want to know that the original time frame was adhered to, and if not, why not. (Concerns have arisen in the past, for example, about authors who only report results from an intermediate time point, and avoid reporting the results from the later time points.) The verification function of a registry therefore requires administrative data, protocol information, and event log.

**4. Systematic Review.** The use of evidence as the foundation of clinical and policy decisions (i.e., evidence based medicine) requires that analysts identify and analyze all of the clinical studies that have been conducted on a given topic and that meet certain criteria (e.g., all randomized double blind controlled trials). This identification process is labor- and time-intensive and is frequently hampered by the fact that not all trials that are conducted are reported in a publicly accessible place. Clinical trials registries have been advocated as a means for identifying a complete list of all relevant studies. Although the data elements necessary to fulfill the "existence" function would help (by identifying the total number of relevant studies to be reviewed), some believe that it would be most helpful to also include more details of the protocol as well as the results.

**5. Public Access.** Access to trial information and results by members of the general public is considered an important societal “good” and is also considered by many to be ethically mandatory given the reliance of these trials on human volunteers. In order to serve this function, the registry would need data elements that cover administrative, protocol information and results.

## **B. Data Elements**

Appendix A lists the specific data elements in ClinicalTrials.gov<sup>4</sup>. Protocol information contains structured fields as well as free text fields with suggested formatting and content. These data elements are currently under review and may be revised to provide more structure and to make them conform to evolving international standards.

Event log information presents challenges for our data providers and for our staff. Some studies are conducted at hundreds of sites. It is therefore difficult for our data providers to keep the contact information and recruitment status up to date. We are exploring ways of working with our data providers to ensure the timeliness of recruitment status and contact information so that potential subjects are not misled.

The results section currently allows for three options. The first and most frequently used option is to provide citations and links (when available) to published articles about results. A second option that we are beginning to explore is to provide a link to an FDA web site, Drugs@FDA<sup>5</sup>, which contains detailed statistical analyses of trials organized by specific drug (e.g., the FDA’s review of clinical trial data submitted as part of the New Drug Application for Vioxx<sup>6</sup>). Although this option takes advantage of an already public, detailed analysis of clinical trial data, the information may or may not be relevant to the specific clinical trial in ClinicalTrials.gov. Finally, data providers who want to display unpublished results are able to link out of ClinicalTrials.gov to their own web site. (Users are notified when they click on a link to another web site.)

There is ongoing discussion across NIH about whether or not unpublished results should be reported in the registry. This issue is also being considered by Congress as part of the Fair Access to Clinical Trials Act (FACT Act, 2005). Advocates believe that without strong incentives for requiring the disclosure of all trial results, some information will never be revealed. On the other hand, others point out that the tasks of analyzing, reporting, and accurately interpreting the results from a clinical trial are not simple, and generally are considered to require significant scientific peer review. ClinicalTrials.gov staff continues to explore possible options as we await direction from Congress and from the NIH working group on this issue.

## **C. Types of Trials**

Figure 1 illustrates the range of clinical research studies (horizontal boxes) along with the sponsors (vertical boxes). The range of studies that are considered by the ICMJE statement and FDAMA 113 are indicated on the right.

---

<sup>4</sup> See <http://prsinfo.clinicaltrials.gov/definitions.html> for data element definitions

<sup>5</sup> <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/>

<sup>6</sup> Available at [http://www.fda.gov/cder/foi/nda/99/021042\\_52\\_Vioxx.htm](http://www.fda.gov/cder/foi/nda/99/021042_52_Vioxx.htm)

Standard textbook definitions of the term “clinical trial” usually include the following features:

- studies in which the investigator prospectively assigns the subject to one of two or more intervention arms;
- subjects receive the assigned intervention(s) and then are followed;
- outcomes are assessed at pre-specified time points;

(These trials are considered “interventional” because the investigator “intervenes” in the subject’s care by assigning the subject to a specific course of treatment or diagnosis.)

In practice, there is some variation in how the term “clinical trial” is used by different institutions and different agencies. Some include “observational” studies (in which the investigator “observes” the subject but does not determine what intervention they receive).

Although IND studies tend to meet the definition of an interventional trial, other studies in ClinicalTrials.gov would be considered observational studies (and some may even be considered basic science). The issue of definition is fundamental to determining which trials should be in the registry (and therefore which trials are “missing”). It is also key to being able to characterize trial design with structured data elements. Currently, different policy initiatives are using different definitions.

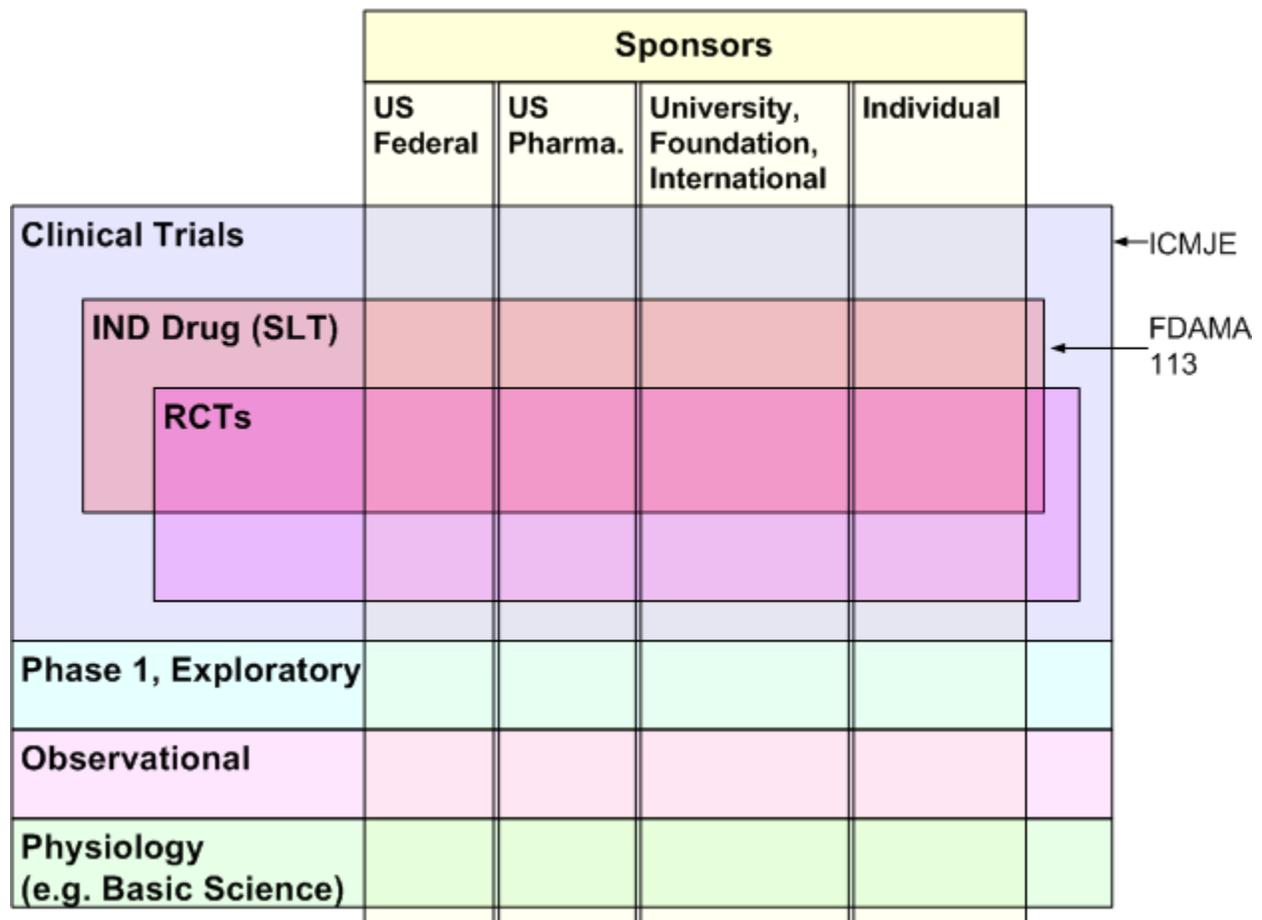


Figure 1. Scope of Clinical Research Studies and Sponsor Types

### III. Policy Landscape

The needs of different user groups (e.g., patients, researchers, policy makers) vary in terms of desired data elements and level of detail. We are exploring options for meeting these different needs.

Figure 2 illustrates ClinicalTrials.gov data elements as required to satisfy various policy initiatives.

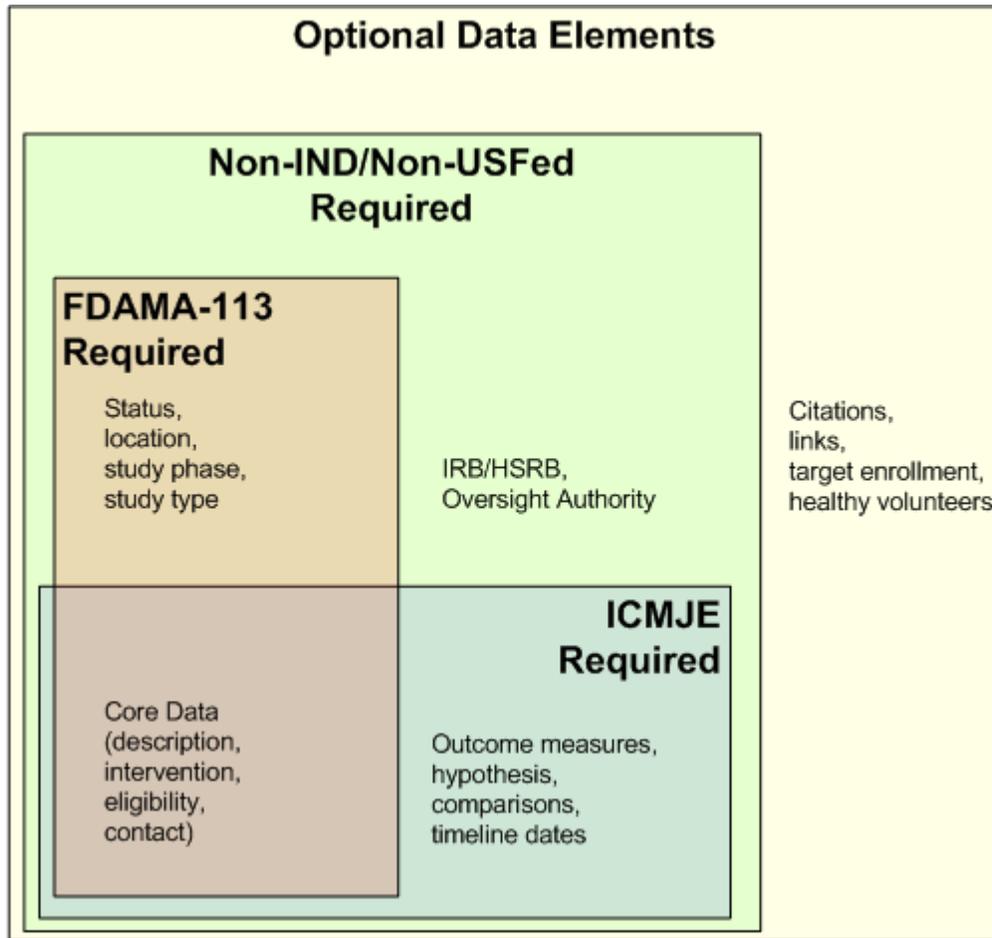


Figure 2. ClinicalTrials.gov Data Elements by Policy Initiative

#### A. Federal Legislation

**1. FDA Modernization Act of 1997, Section 113 (FDAMA 113).** Although this law sets forth specific requirements for trial registration, it has several limitations. First, there is no clear definition of “serious or life threatening” so that the universe of trials that should be registered cannot be identified. Second, the law does not provide a specific enforcement mechanism. Third, the law only applies to drug studies and does not include other medical interventions (e.g.,

devices, surgical procedures). Many of the policy initiatives discussed below are designed to expand the scope of ClinicalTrials.gov as determined by FDAMA 113 by expanding the types of trials and the required data elements.

**2. Best Pharmaceuticals for Children Act of 2002 (BPCA).** The BPCA requirements call for reporting of information related to “compassionate use” access of investigational new drugs (e.g., protocol exceptions, single-patient, expanded access protocols), especially for children. “The BPCA, signed by the President on January 4, 2002, requires a description of whether, and through what procedure, the manufacturer or sponsor of an IND will respond to requests for protocol exception, with appropriate safeguards, for single-patient and expanded access use of the investigational drug, particularly in children” (FDA, 2004). The FDA issued a Draft Guidance for Industry in January 2004 incorporating BCPA data collection requirements for ClinicalTrials.gov. Simultaneous with the release of the Draft FDA Guidance, ClinicalTrials.gov provided a “mock up” PRS mechanism for collecting these data<sup>7</sup>. A final FDA Guidance Document and activation of the PRS implementation is anticipated in 2005.

**3. Fair Access to Clinical Trials Act of 2005 (FACT).** The Fair Access to Clinical Trials (FACT) Act was introduced in the Senate in February 2005 (S. 470) in response to several recent well-publicized events related to unreported clinical trial results. The bill would expand FDAMA 113 to provide for enforcement and to require registration of device studies in ClinicalTrials.gov. In addition, the FACT Act would also mandate the reporting of all clinical trial results, positive and negative. Passage of the FACT Act would have a significant impact on the operations and scope of ClinicalTrials.gov.

## **B. International Committee of Medical Journal Editors (ICJME) Statement**

In September 2004, the International Committee of Medical Journal Editors (ICJME) issued a statement regarding clinical trial registration (DeAngelis, 2004). The ICMJE announced a new policy to reduce potential misrepresentation of clinical trial results in publications. The policy calls for the mandatory registration of clinical trials before the first patient is enrolled as a condition for consideration for publication. Further, the statement specifies particular data that must be provided by the sponsor to an acceptable registry.

In conjunction with publication of the ICMJE statement, ClinicalTrials.gov created new data fields and modified others to accommodate these criteria (see Figure 3). In addition, while only submission of trials conducted under IND or funded by the NIH and other Federal agencies were permitted previously, the NLM agreed “to accept validated descriptions of all clinical trials without charge from the international community for inclusion in ClinicalTrials.gov” (ICMJE, 2004).

---

<sup>7</sup> <http://prsinfo.clinicaltrials.gov/ind-bpca.html>

<b>ICJME Criterion</b>	<b>ClinicalTrials.gov Data Element</b>	<b>ClinicalTrials.gov Display</b>
Unique Identifying Number	[Assigned by System]	ClinicalTrials.gov Identifier
Statement of the Intervention(s)	Intervention Name	Treatment or Intervention
Comparison(s) Studied	Detailed Description	Further Study Details
Statement of the Study Hypothesis	Detailed Description	Further Study Details
Primary Outcome Measures	Primary Outcomes	Primary Outcomes
Secondary Outcome Measures	Secondary Outcomes	Secondary Outcomes
Eligibility Criteria	Eligibility Criteria	Eligibility
Key Trial Dates		
Registration Date	[Assigned by System]	Record first received
Start Date	Study Start Date	Study Start
Date of Last Follow-up	Last Follow-Up Date	Last Follow-Up
Date of Closure to Data Entry	Data Entry Closure Date	Data Entry Closure
Date Trial Data Considered Complete	Study Completion Date	Study Completion
Target Number of Subjects	Target Number of Subjects	Expected Total Enrollment
Funding Source	Sponsor & Collaborators	Sponsored by
Investigator Contact Information	Locations: Investigator Name Study Official's Name	Location and Contact Information

Figure 3. Mapping of ICMJE Criteria to ClinicalTrials.gov (2/10/05)

### **C. World Health Organization (WHO)**

The World Health Organization (WHO) recognizes that comprehensive registration of clinical trials is important for improving global public health (Evans, Gülmezoglu, & Pang, 2004). WHO has established several working groups to explore various issues related to organizing international trial registries, such as assigning of unique identifiers and developing a single portal to provide access to registries around the world.<sup>8</sup>

### **D. Pharmaceutical Industry**

In January 2005, the International Federation of Pharmaceutical Manufacturers and Associations (IFPMA) and three regional trade associations, including the Pharmaceutical Research and Manufacturers of America (PhRMA), issued a Joint Position on the Disclosure of Clinical Trial Information<sup>9</sup>. Under the statement, the pharmaceutical companies recommended voluntary submission of information about all non-exploratory industry-sponsored trials to publicly accessible registries, such as ClinicalTrials.gov. Furthermore, the statement suggests that all members submit summary results of industry-sponsored trials, regardless of outcome, for drugs that have received marketing approval.

<sup>8</sup> Dr. Zarin will be attending a meeting at the World Health Organization in April and will be able to report on any new developments during the May Board of Scientific Counselors meeting.

<sup>9</sup> [http://www.ifpma.org/Documents/NR2205/joint%20position\\_clinical%20trials.PDF](http://www.ifpma.org/Documents/NR2205/joint%20position_clinical%20trials.PDF)

## E. State Legislation

In late 2004, Consumers Union initiated a grassroots campaign called, “Prescription for Change.”<sup>10</sup> The campaign urges citizens to email members of the Federal and state legislatures to support issues related to clinical trials and drug safety. As of April 2005, bills have been introduced into nine state legislatures (i.e., CA, CT, MD, MN, NJ, SC, TN, TX, and VT) calling for registration of clinical trials that are conducted in that state and, in many cases, the release of results.

## F. Cochrane Collaboration

The Cochrane Collaboration<sup>11</sup>, an international not-for-profit organization established in 1993, creates and distributes “systematic reviews of healthcare interventions and promotes the search for evidence in the form of clinical trials and other studies of interventions.” In October 2004 at the 12th Cochrane Colloquium, a working group drafted the first part of the “Ottawa Statement,” which argues for international trial registration based on scientific and ethical principles. The document has been submitted for publication.

In summary, the overall policy landscape includes a range of data elements that are required to meet the needs of various user groups. As shown in Figure 4, ClinicalTrials.gov will potentially need to continually adapt in an effort to address each of these needs.

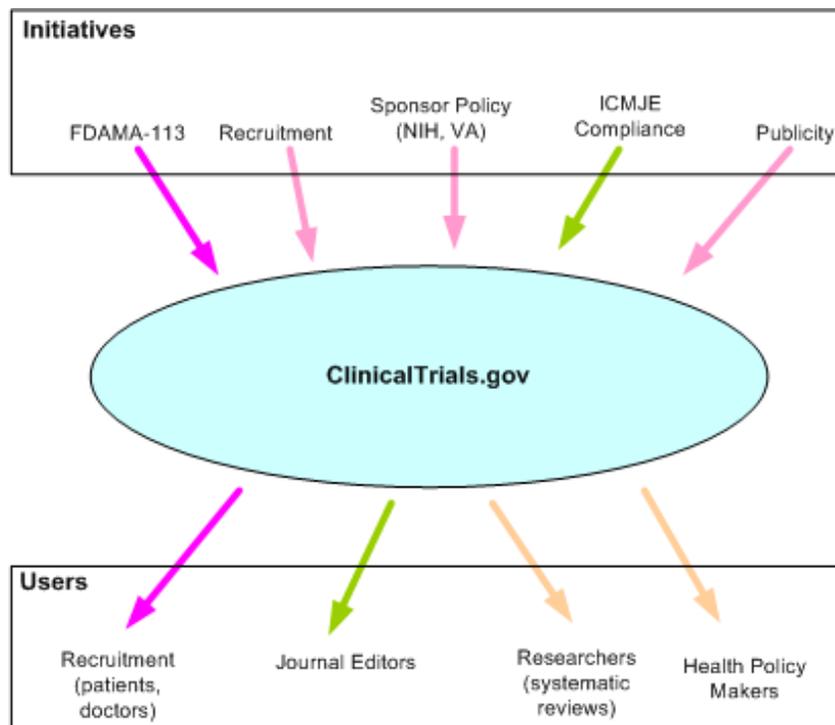


Figure 4. Motivation for Registration with ClinicalTrials.gov

<sup>10</sup> <https://secure2.convio.net/cu/site/Advocacy?page=UserAction&cmd=display&id=357>

<sup>11</sup> <http://www.cochrane.org/>

## IV. Data Quality Issues

### A. The Denominator Problem: Definition and Compliance

One significant justification for trial registration is to monitor and track all studies in existence. The “denominator problem” refers to the fact that the total number of clinical trials conducted and that *should* be registered, is not known. Without a denominator, it is not possible to determine the completeness of a trial registry such as ClinicalTrials.gov. Further, a subtle, but related complexity is defining a standard unit of analysis in counting number of trials. Some registries, such as CenterWatch<sup>12</sup>, consider each individual site of a multi-site study to be a different trial. Others, such as ClinicalTrials.gov, consider all sites conducting a protocol to be a single trial. Thus, comparisons of numbers of trials between registries must take into account the unit of analysis.

The concept of “compliance” can be considered for those trials that fall under the scope of FDAMA 113 since their sponsors are legally required to register them in ClinicalTrials.gov. In addition, we have encouraged the registration of other trials funded by NIH, though there is no legal requirement and no formal policy at this time. (Although other trials may be registered for a variety of reasons, including the ICMJE requirement, the concept of “compliance” does not apply to those trials.)

ClinicalTrials.gov does not have a mechanism for determining, with precision, the degree to which investigators and sponsors are complying with FDAMA 113. First, the scope of this legislation is not clear since the phrase “serious or life threatening diseases and conditions” is subject to interpretation (FDA, 2002). Even for trials that are clearly “within scope” (e.g., cancer and AIDS trials) ClinicalTrials.gov is dependent on the FDA to determine the full list of relevant trials. The FDA conducted a small evaluation study of compliance with FDAMA 113 using a sample of cancer trials. The final report has not yet been released, but the preliminary report indicates that “48% of the mandated industry-sponsored and 91% of the mandated NIH cancer-related trials were in ClinicalTrials.gov” (Toigo, 2004). Other than this one evaluation project, FDA has not been able to provide a “denominator” in the context of FDAMA 113.

ClinicalTrials.gov is also hampered in its ability to assess the percentage of NIH clinical trials included in the registry. The lack of a formal NIH policy leaves ambiguity about which trials should be registered. Many definitions of “clinical trials” are used across NIH institutes and centers, and they each have their own preferences regarding registration. Furthermore, there is no centralized database other than ClinicalTrials.gov that includes information about NIH-funded clinical trials. However, as part of work done under the NIH Roadmap, new grants are now being coded by whether or not they include a clinical trial. In addition, tracking of subjects within clinical trials is also being done in a way that may provide some mechanisms for identifying NIH sponsored clinical trials over the course of the next several years. For now, ClinicalTrials.gov works with liaisons at each NIH institute and center to facilitate registration of its clinical trials.

---

<sup>12</sup> <http://www.centerwatch.com/>

## B. The Verification Problem

ClinicalTrials.gov contains information about clinical trials that is provided to us by “data providers.” Most of the information provided is not publicly available in other forms. As a result, independent verification of most of the data is not possible.

In some instances, additional information could help to verify key characteristics of the trial. However, FDAMA 113 limits our ability to require additional information. ClinicalTrials.gov currently requires additional data elements for “non-FDAMA” trials in an effort to provide more confidence in the accuracy of the data (Figure 2).

## C. Current Quality Assurance Procedures

ClinicalTrials.gov uses a distributed data model. Clinical trial sponsors or their designated surrogates (collectively known as “data providers”) submit summary protocol information through the Web-based Protocol Registration System (PRS) using one of two mechanisms.

Currently, quality assurance (QA) efforts at ClinicalTrials.gov focus on ensuring internal consistency and logic of entries, as well as ensuring that all links are active. Data quality controls at the PRS interface include data validation rules, use of controlled vocabularies (e.g., MeSH), and workflow control through the PRS user and administrator roles (Figure 5).

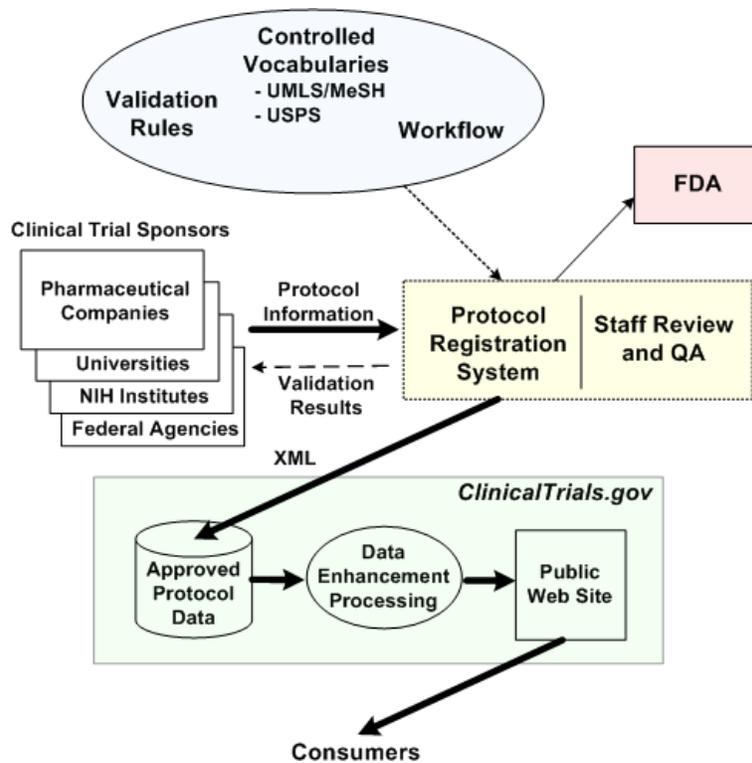


Figure 5. ClinicalTrials.gov Data Flow

Once a data entry for a record is completed, the ClinicalTrials.gov staff manually review the data using several reporting tools. Minor editorial issues (e.g., typos) are corrected by the staff and noted in the comment field. Substantive issues (e.g., inconsistencies in content) are forwarded to the data provider for resolution. After a record is approved by staff, it is processed and enhanced (e.g., linked to relevant MedlinePlus health topic pages) and published to the public site.

ClinicalTrials.gov staff reviews all new and modified records released by data providers daily using an online report through the PRS. New records are reviewed for typos, consistency, and glaring content errors and modified records are reviewed for general “correctness.” The ClinicalTrials.gov staff contact up to three locations, selected at random, listed in all recruiting records to confirm the accuracy and currency of the location information. New and modified records that appear to be error-free are published. Records for which errors are detected are withheld for further analysis and possible action.

The staff scrutinizes specific aspects of published records. The following activities are conducted at regular intervals:

- Review new hyperlinks entered into ClinicalTrials.gov records
- Spell check text fields
- Identify potential duplication of records
- Check for PubMed citations

In addition to these QA/QC activities the PRS automatically generates a report that lists active protocol records that have not been updated in the past 6 months and active protocol records that have not been verified in the past 6 months. When data providers log into their PRS accounts, they are reminded that these records need to be reviewed.

## **V. Implementation**

ClinicalTrials.gov has been designed with an emphasis on ease of use and providing relevant medical information for non-health professionals, including patients, family members, and other health consumers. Visitors may access comprehensive, accurate, and timely information on clinical trials within a few mouse clicks. All study data are presented in a standard format with four sections: Purpose, Eligibility, Locations and Contact Information, and More Information. The condition, treatment or intervention, and phase of each study are highlighted in a table for easy reference.

Additional context-sensitive resources for non-professionals are provided through numerous “just-in-time” hyperlinks to relevant documents in other NIH online biomedical resources. Thus, visitors who want to learn more about a condition being studied or the published biomedical literature related to a study simply click on hotlinks provided in ClinicalTrials.gov records. From within a record, they may access related health topics in MedlinePlus, NIH’s consumer health resource, and specific citations in Pub-Med/MEDLINE, NIH’s bibliographic database (Figure 6). The overall design of ClinicalTrials.gov emphasizes usability, accessibility (e.g., compliance with Section 508 of the ADA), and contextualization (e.g., just-in-time hotlinks). The site relies on natural language processing tools, terminological systems, and other medical informatics tools developed at NIH.

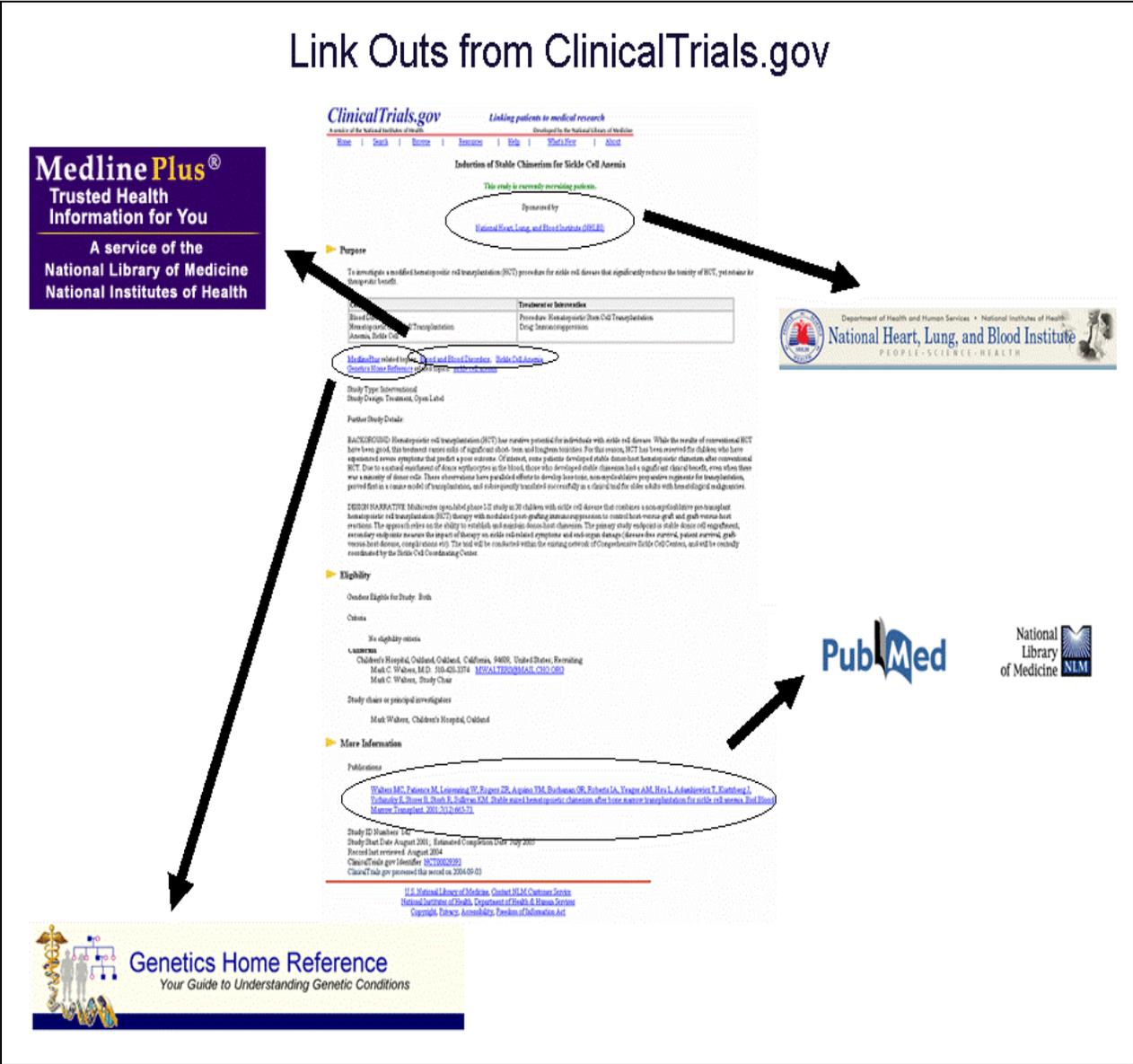


Figure 6. Context-Sensitive Links to Just-in-Time Health Information

**A. System Design**

The implementation approach for ClinicalTrials.gov is to collect trial records from the sponsors of the trials into a centralized database at NLM. The data are entered directly into a web-based data entry system by the sponsors. After the data have been reviewed by ClinicalTrials.gov staff, the data are then processed and made available to the public. Figure 7 shows the overall system data flow from the content providers through data capture and preparation and finally to the public web site.

## ClinicalTrials.gov: System Data Flow

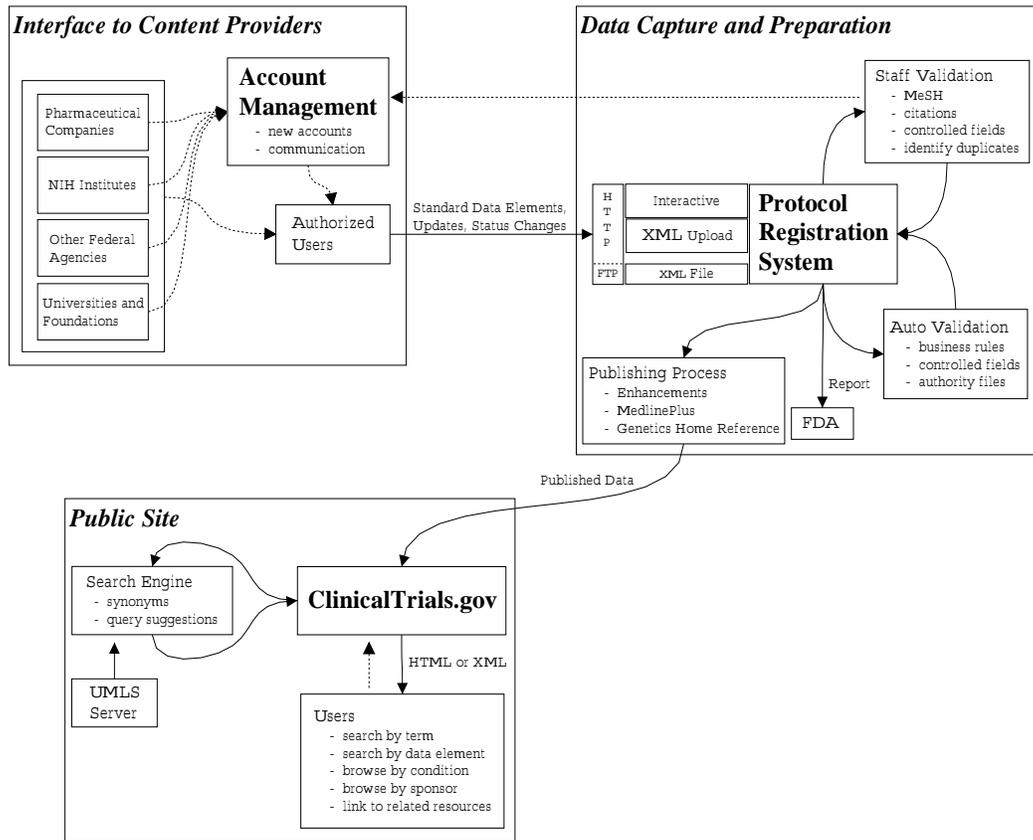


Figure 7. Diagram of System Components

The Protocol Registration System is a web based data entry system developed at LHCNBC to facilitate the collection of the trial data from sponsors (Gillen et al, 2004). In essence, the system provides forms to capture the data elements required by ClinicalTrials.gov. Business rules in the Protocol Registration System regulate which fields are required and which are optional. In addition, the system provides tools to facilitate entry of controlled vocabulary and provides various data integrity checks.

The publishing process shown in Figure 7 was first described in McCray and Ide (2000). The data preparation part of the system has not changed significantly since that time. Additional data elements have been added and some of the processing steps have been optimized, but conceptually the process is the same. In summary, the study records in XML form are deposited into a directory for processing. Every night the data are copied from the source repository and processed. The processing includes validating the data elements for consistency and adherence to standards and enhancing the data with links to additional resources, such as MedlinePlus, Genetics Home Reference (GHR)<sup>13</sup>, and PubMed. The data are then indexed by the search system and distributed to the public web servers.

<sup>13</sup> <http://ghr.nlm.nih.gov/>

More recently, the search system underlying ClinicalTrials.gov was completely replaced by a system developed by the ClinicalTrials.gov technical team (McCray et al., 2004). In addition, this search engine has been applied in the Genomics Track of TREC 2003 and TREC 2004 (Kaylaap et al, 2003).

The search system takes advantage of the underlying document structure to adjust the ranking of search results. In addition, the system allows for lexical variation, synonyms, and flexible query interpretation. In broad strokes, a document that has the “right words” in the “right places” is considered to be good and receives a high rank. The “right words” are the user’s search terms, their variants, and their synonyms. The “right places” in descending order of importance are the Study Title, Condition, Intervention, Brief Summary, and the Location fields.

## **B. Research and Development**

LHNCBC has been an ideal home for ClinicalTrials.gov. The language and medical resources maintained by NLM/LHNCBC have been of tremendous utility in building an effective consumer website. In addition, these resources have been enriched through focusing on the practical requirements imposed by their use in ClinicalTrials.gov. In addition, the ClinicalTrials.gov project has created a number of tools which have been employed by other projects at LHNCBC. Thus, as shown in Figure 8, ClinicalTrials.gov is both a “consumer” of NLM/LHNCBC resources and also a “producer” of new services.

**1. “Consumer” of NLM Products.** ClinicalTrials.gov uses the following tools developed at NLM/LHNCBC:

- Medical Subjects Heading (MeSH) Vocabulary
- MEDLINE/PubMed – biomedical bibliographic citation database
- MedlinePlus – consumer health information resource
- Unified Medical Language System (UMLS) and the Terminology Server
- Restrict to MeSH
- SPECIALIST Lexicon
- G-spell: spell checking and recommendation
- Lexical Tools: MMTx

MeSH terms are used in the condition field of studies whenever possible. The publication processing uses MeSH as a mapping mechanism to identify possibly relevant health topics from MedlinePlus and then the studies are annotated with links to those topics. The same mechanism is used to link to topics in Genetics Home Reference. In some cases, the terms provided by the study sponsors are not MeSH terms. In those cases, the terms are mapped to the UMLS through the Knowledge Source Server and then the Restrict-to-MeSH table created by Dr. Olivier Bodenreider is used to map these terms to MeSH.

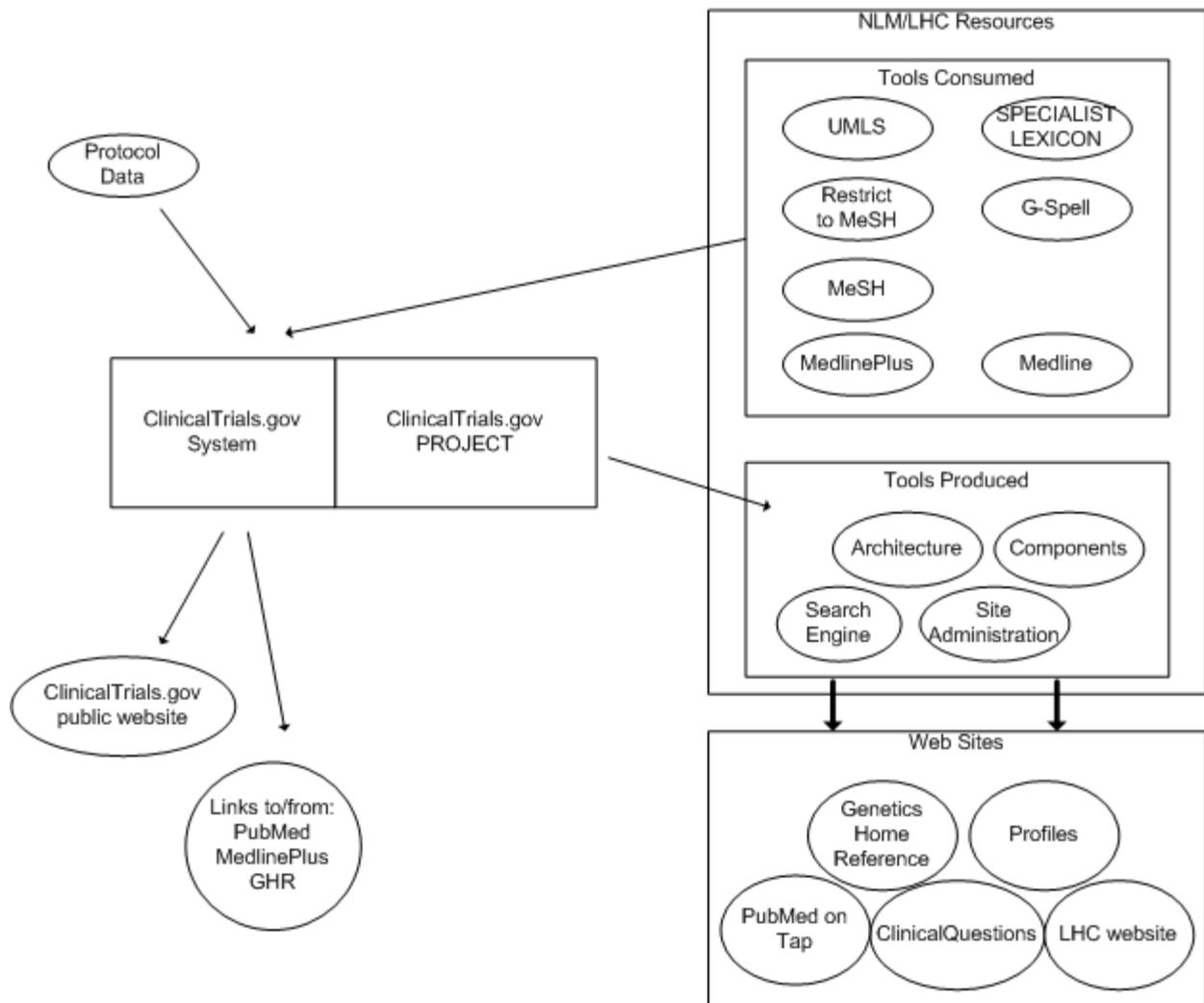


Figure 8. ClinicalTrials.gov Resources and Products

Studies may also contain references to the published literature. These citations are mapped to PubMed identifiers and the studies enhanced to contain links to the appropriate abstracts in PubMed.

The lexical tools, specialist lexicon, and UMLS are used by the search system to enhance the user's query terms with variants and synonyms to increase search recall. In addition, the G-spell tool developed by the lexical systems group at LHNCBC is used to provide the user with spelling suggestions.

The specialist lexicon is also used to assist the ClinicalTrials.gov quality assurance team in identifying spelling errors in the study records.

**2. “Producer” of NLM Products.** The ClinicalTrials.gov project has spawned a number of products and tools that contribute to the resources available at LHNCBC.

- Protocol Registration System framework
  - The software developed for the Protocol Registration System data entry tool has served as a model for similar systems. Much of the underlying framework has been used by the Genetics Home Reference project and the ClinicalQuestions project.
- Site administration tools
  - A number of administrative tools have been developed during the course of the ClinicalTrials.gov project. These tools assist the team with site maintenance and monitoring. Several other projects have begun using these tools, including Genetics Home Reference and the new implementation of Profiles in Science<sup>14</sup>.
- Cross-language Information Retrieval Spanish Components
  - A Spanish version of ClinicalTrials.gov has been under development for the last year. This project has involved some research into cross-language information retrieval (Roseblat et al., 2004; submitted). As a result of this project, we have constructed a Spanish-English bilingual term list which may have benefits to other projects.
- SE Search Engine
  - The search engine that has been developed by the ClinicalTrials.gov team has proven to be effective and readily adaptable to other projects. Other projects utilizing the search engine are:
    - Genetics Home Reference
    - Profiles in Science
    - ClinicalQuestions
    - PubMed OnTap
    - LHNCBC Internal Website
    - SPECIALIST Lexicon: The curators of the lexicon have been using the SE search engine and the search engine’s user interface to mine MEDLINE for terminology and word usage. In addition, a number of useful reports have been generated as a by-product of the search engine’s indexing of MEDLINE.

## **VI. Future Directions/Challenges**

ClinicalTrials.gov must continue to adapt to the changing policy environment. Depending on decisions made by ICMJE, WHO, The Congress of the United States and various State governments, ClinicalTrials.gov may need to expand and/or modify some of its data elements. ClinicalTrials.gov may also need to collect information on unpublished results, which would require substantial study prior to implementation. As these events develop, ClinicalTrials.gov will focus on the following areas:

---

<sup>14</sup> <http://profiles.nlm.nih.gov/>

1. Increasing registration rates of clinical trials, in general, and improving compliance with any laws that mandate registration;
2. Evaluating the needs and experiences of different user groups and modifying the site as necessary to better meet those needs; and
3. Improving our procedures for data verification and modifying data element definitions as necessary.

The following three initiatives illustrate our current approach to meeting some of these challenges as we await further policy developments:

#### **A. Improve Registration Rates of Relevant Trials**

Although we cannot precisely measure the compliance rates for either the pharmaceutical company trials that are under FDAMA 113 or for NIH institutes/centers, we know that the rates are not 100%. Our efforts to improve these rates include:

1. Work with FDA to obtain permission for our staff to have access to the IND records so that we can conduct our own compliance studies;
2. Work with individual pharmaceutical companies to improve procedures for ensuring registrations of all mandated trials;
3. Work with individual institutes and centers at the NIH to develop and implement procedures for increasing clinical trial registration;
4. Work with the Office of the Director at NIH to clarify and communicate NIH policy in trial registration;
5. Work with the Office of Extramural Research at NIH to develop data monitoring strategies across NIH so that we will have access to “denominator” data;
6. Work with individual institute and center staff, as well as investigators, to improve methods to ensuring timely updates of existing ClinicalTrials.gov records.

#### **B. Evaluate the Utility and Feasibility of Reference Protocols**

Sponsors who register trials outside of the scope of FDAMA 113 face stricter registration requirements that are designed to provide us with more detailed information about the protocol. Some sponsors have objected to these more stringent requirements because they believe that it forces them to reveal “proprietary information” at a date that is earlier than necessary to meet the public policy objectives of authenticating data upon submission to a journal. In addition, we have no mechanism to verify these data elements. In order to resolve both of these issues, we have suggested a program in which willing sponsors will submit a full protocol (e.g., as a PDF file) along with their initial registration.. This “reference copy” would only be available to NLM staff for conducting QA/QC activities on the validity of the information provided to ClinicalTrials.gov. By doing so, they may be able to meet ICMJE requirements and still wait to enter some of the data elements until a later date. (If and when a paper is submitted for publication, a journal editor could “authenticate” information provided by the authors against the information in the original protocol maintained at NLM.) Such a mechanism has several precedents, including NLM GenBank holding sequence submissions prior to publication and

NCI holding and extracting protocol information for Cancer.gov records. Issues related to the Freedom of Information Act (FOIA) remain unresolved.

Despite some of the uncertainties, we have two pharmaceutical companies which have expressed a willingness to work with us on a pilot study. We are developing a plan for testing the feasibility and utility of collecting protocols and using them to verify registration information.

### **C. Pilot Study of Clinical Trial Verification Procedures**

We are in the process of testing verification procedures for the following four types of trials in ClinicalTrials.gov:

1. Trials funded by NIH;
2. Trials sponsored by a pharmaceutical company and being submitted under FDAMA 113
3. Unaffiliated domestic trials
4. Unaffiliated international trials

Although there has been the assumption that trials in categories a) and b) are “verified” by our Federal partners, we are evaluating the validity of data in all of these categories in a small pilot study. Nine trials (the maximum number that does not require specific OMB clearance) from each category are being randomly chosen. For each category, we are developing a short questionnaire. We will contact either the Federal partner (for a and b), the Health Authority, and/or the IRB. We will attempt to verify the existence of the trial as well as determine the ability of our contact to verify specific data elements. The results will inform future decisions about QA/QC procedures.

## **VII. References**

- DeAngelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Kotzin S, Laine C, Marusic A, Overbeke AJ, Schroeder TV, Sox HC, Van Der Weyden MB; International Committee of Medical Journal Editors. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med*. 2004 Sep 16;351(12):1250-1. Epub 2004 Sep 08.
- Dickersin K, Rennie D. Registering clinical trials. *JAMA*. 2003 Jul 23;290(4):516-23.
- Evans T, Gulmezoglu M, Pang T. Registering clinical trials: an essential role for WHO. *Lancet*. 2004 May 1;363(9419):1413-4.
- FACT Act (Fair Access to Clinical Trials Act of 2005). 2005. S.470. Introduced in the U.S. Senate on February 28, 2005 by Senators Dodd, Grassley, Johnson, and Wyden.
- FDA. *Guidance for Industry: Information Program on Clinical Trials for Serious or Life-Threatening Diseases and Conditions*. Rockville, MD: U.S. Food and Drug Administration. 2002, March. Available at: <http://www.fda.gov/cder/guidance/4856FNL.PDF>
- FDA. *Draft Guidance for Industry: Information Program on Clinical Trials for Serious or Life-Threatening Diseases and Conditions*. Rockville, MD: U.S. Food and Drug Administration, Revision 1. 2004, January. Available at: <http://www.fda.gov/OHRMS/DOCKETS/98fr/2004d-0014-gdl0001.pdf>

- FDAMA 113. *Food and Drug Modernization Act, Section 113*. 1997.  
<http://www.fda.gov/cder/guidance/105-115.htm#SEC.%20113>
- Gemoets D, Rosemblat G, Tse T, Logan R. Assessing readability of consumer health information: an exploratory study. *Medinfo*. 2004;2004:869-73.
- Gillen JE, Tse T, Ide NC, McCray AT. Design, implementation and management of a web-based data entry system for ClinicalTrials.gov. *Medinfo*. 2004;2004:1466-70.
- ICMJE. *Update on Trials Registration: October 2004*. Available at:  
[http://www.icmje.org/icmje\\_response.pdf](http://www.icmje.org/icmje_response.pdf)
- Kayaalp M, Aronson AR, Humphrey SM, Ide NC, Tanabe LK. Methods for accurate retrieval of MEDLINE citations in functional genomics. *Proceedings of the Text Retrieval Conference (TREC-2003)*. 2003:175-84.
- McCray AT, Ide NC, Loane RR, Tse T. Strategies for supporting consumer health information seeking. *Medinfo*. 2004;2004:1152-6.
- McCray AT, Dorfman E, Ripple A, Ide NC, Jha M, Katz DG, Loane RF, Tse T. Usability issues in developing a Web-based consumer health site. *Proc AMIA Symp* 2000:556-60.
- McCray AT, Ide NC. Design and implementation of a national clinical trials registry. *J Am Med Inform Assoc*. 2000 May-Jun;7(3):313-23.
- Rosemblat G, Tse T, Gemoets D, Gillen JE, Ide NC. Supporting access to consumer health information across languages. Submitted to *AMIA 05*.
- Rosemblat G, Tse T, Gemoets D. Adapting a monolingual consumer health system for Spanish cross-language information retrieval. *ISKO*. 2004;9:315-21.
- Toigo T. Food and Drug Modernization Act (FDAMA) Section 113: status report on implementation. *Biolaw and Business*. 2004;7(2). Available at:  
<http://www.fda.gov/oashi/clinicaltrials/section113/biolaw.html>.
- Tse T, Logan RA. Towards a more comprehensive conceptual framework for consumer health information seeking. Submitted to *AMIA 05*.
- Tse T, Johnson SO, Ripple A. Update: Expanded ClinicalTrials.gov Search Capabilities. *NLM Tech Bull*. 2002 Jul-Aug;(327):e5. Available at:  
[http://www.nlm.nih.gov/pubs/techbull/ja02/ja02\\_ctgov.html](http://www.nlm.nih.gov/pubs/techbull/ja02/ja02_ctgov.html).

**Appendix A**  
**Data Elements for ClinicalTrials.gov**

<p><b>1. Titles and Background Information:</b></p> <ul style="list-style-type: none"><li>• Organization's Unique Protocol ID</li><li>• Secondary IDs (maximum of 5)</li><li>• Brief Title</li><li>• Official Title</li></ul>
<p><b>2. Investigational New Drug Application (IND)/Investigational Device Exemption (IDE) Information:</b> <i>Will not be made public - for administrative purposes only</i></p> <ul style="list-style-type: none"><li>• IND/IDE Protocol? (yes/no)</li><li>• IND/IDE Grantor (CDER/CBER)</li><li>• IND/IDE Number (or "Not Yet Assigned")</li><li>• IND/IDE Serial Number</li></ul>
<p><b>3. Human Subjects Review</b></p> <ul style="list-style-type: none"><li>• Board Approved?</li><li>• Board Approval Number</li><li>• Board Name</li><li>• Board Affiliation</li><li>• Board Chair<ul style="list-style-type: none"><li>○ Name</li><li>○ Phone</li><li>○ Extension</li><li>○ Email</li><li>○ Address</li></ul></li><li>• Oversight Authorities</li></ul>
<p><b>4. Sponsors:</b></p> <ul style="list-style-type: none"><li>• Sponsor</li><li>• Collaborators (maximum of 10)</li></ul>
<p><b>5. Study Description:</b></p> <ul style="list-style-type: none"><li>• Brief Summary</li><li>• Detailed Description</li></ul>
<p><b>6. Status:</b></p> <ul style="list-style-type: none"><li>• Study Phase<ul style="list-style-type: none"><li>○ Phase 1</li></ul></li></ul>

- Phase 1/Phase 2
- Phase 2
- Phase 2/Phase 3
- Phase 3
- Phase 4
- N/A (Used rarely. IND studies must not use this option)
- Study Type (Interventional/Observational)
- Overall Recruitment Status
  - Not yet recruiting
  - Recruiting
  - No longer recruiting
  - Completed
  - Suspended
  - Terminated
- Record Verification Date
- Start Date
- Last Follow-Up Date
- Date Entry Closure Date
- Completion Date

**7a. Study Design (Study Type: Interventional):**

- Purpose
  - Treatment
  - Prevention
  - Diagnosis
  - Educational/Counseling/Training
- Allocation
  - Randomized Controlled Trial
  - Non-randomized Trial
- Masking
  - Open
  - Single Blind
  - Double Blind
- Control
  - Placebo
  - Active
  - None
  - Historical
  - Dose Comparison
- Assignment
  - Single Group
  - Parallel
  - Cross-over
  - Factorial
  - Expanded Access

- Endpoint
  - Safety
  - Efficacy
  - Safety/Efficacy
  - Bio-equivalence
  - Bio-availability
  - Pharmacokinetics
  - Pharmacodynamics
  - Pharmacokinetics/Pharmacodynamics
- Outcomes
  - Primary Outcomes
  - Secondary Outcomes

**7b. Study Design (Study Type: Observational):**

- Purpose
  - Natural History
  - Screening
  - Psychosocial
- Duration
  - Longitudinal
  - Cross-sectional
- Selection
  - Convenience Sample
  - Defined Population
  - Random Sample
  - Case Control
- Timing
  - Retrospective
  - Prospective
  - Both

**8. Interventions (maximum of 10):**

- Intervention Type
  - Drug
  - Gene Transfer
  - Vaccine
  - Behavior
  - Device
  - Procedure
- Intervention Name

**9. Conditions and Keywords:**

- Conditions (maximum of 5)

- Keywords

**10. Eligibility:**

- Eligibility Criteria
- Gender (Both/Female/Male)
- Minimum Age
- Maximum Age
- Accepts Healthy Volunteers? (yes/no)
- Target Number of Subjects

**11. Protocol Location, Contact and Investigator Information:**

- Facility
  - Name
  - City
  - State/Province
  - Postal Code
  - Country
- Recruitment Status
- Facility Contact
  - First Name
  - Middle Initial
  - Last Name
  - Degree
  - Phone
  - Ext
  - Email
- Facility Contact Backup
- Investigators
  - First Name
  - Middle Initial
  - Last Name
  - Degrees
  - Role (Principal Investigator/Sub-Investigator)
- Central Contact
  - First Name
  - Middle Initial
  - Last Name
  - Degree
  - Phone
  - Ext
  - Email
- Central Contact Backup
- Overall Study Officials
  - First Name

- Middle Initial
- Last Name
- Degree
- Official's Role (Study Chair/Study Director/Principal Investigator)
- Organizational Affiliation

**12. Related Information:**

- References:
  - MEDLINE Identifier
  - Citation
  - Results Reference?
- Links:
  - URL
  - Description

## **Appendix B**

### **Questions for the Board**

1. Given the number of distinct stakeholder groups (e.g., patients, physicians, researchers, systematic reviewers, and policy makers), what evaluation methods/techniques does the Board suggest for (1) identifying the specific needs of each group and (2) assessing how well ClinicalTrials.gov is meeting them?
2. Does the Board have any recommendations for evaluating the feasibility/effectiveness of collecting and using reference copies of trial protocols for ClinicalTrials.gov quality assurance activities?
3. Are there particular informatics techniques that the Board feels would be relevant and useful to ClinicalTrials.gov?

## ***Deborah A. Zarin, M.D.***

Director, ClinicalTrials.gov &  
Assistant Director for Clinical Research Projects

### **Education and Training**

Stanford University	B.A.	1977
Harvard Medical School	M.D.	1981

### **Recent Professional Experience**

#### ***Employment***

Director, ClinicalTrials.gov & Assistant Director for Clinical Research Projects Lister Hill National Center for Biomedical Communications, National Library of Medicine, NIH, DHHS	1/2005 - present
Director, Technology Assessment Program Agency for Healthcare Research and Quality (AHRQ), DHHS	2/2000 – 1/2005
Deputy Medical Director, American Psychiatric Association (APA)	2/1995 – 1/2000
Director, Office of Quality Improvement and Psychiatric Services, APA	7/1998 – 1/2000
Co-Director, Office of Research, & Director, Women's And Children's Affairs, APA	8/1995 – 7/1998
Associate Director, Office of Research. APA	8/1992 – 8/1995

#### ***Training***

Internship in Pediatrics, Massachusetts General Hospital, Boston	7/1981 – 6/1982
Fellowship in Clinical Decision Making, New England Medical Center, Boston, MA	7/1982 – 6/1983
Residency in Psychiatry, McLean Hospital, Belmont, MA	7/1983 – 6/1985
Fellowship in Child and Adolescent Psychiatry Bradley Hospital, Brown University, Providence, RI	7/1985 – 9/1987

#### ***Professionals Certifications***

General Psychiatry, American Board of Psychiatry and Neurology, 1987  
Child and Adolescent Psychiatry, American Board of Psychiatry and Neurology, 1988

#### ***Honors and Awards***

American Academy of Child & Adolescent Psychiatry Elaine Schossler Lewis Award for  
Research on Attention-Deficit Disorder, 1999  
AHRQ: Special Act or Service Award for staffing of Clinical Trials Coverage Project, 2000  
AHRQ: Superior Teamwork Award for Technology Assessment Program, 2002  
AHRQ: Superior Teamwork Award for Technology Assessment Program, 2003

#### ***Professional Memberships***

Society for Medical Decision Making	1983-2003
American Psychiatric Association	1985 – 2001
Massachusetts Psychiatric Society	1985 – 1993
New England Council for Child Psychiatry	1987 – 1993
American Academy of Child and Adolescent Psychiatry	1989 – present
Washington Psychiatric Society	1993 – 2001
Association for Health Services Research	1996 – 2003

#### **Research Activities**

Director, Technology Assessment Program, Agency for Healthcare Quality and Research, 2000-2004	
Director, Practice Guidelines Program, American Psychiatric Association	1992 – 1999
Co-Director, Practice Research Network, American Psychiatric Association	1995 – 1998
Co-Principal Investigator, National Psychiatric Practice Research Network, John D. and Catherine T. MacArthur Foundation, Chicago, IL	1996-1998

Project Director, Examination of the Use of Medications for Children and Adolescents with ADHD, Department of Health and Human Services	1997 – 1998
Principal Investigator, Examination of Characteristics and Outcome of Substance Abusing Clients Treated by Members of the American Psychiatric Association, Center for Substance Abuse Treatment	1997
Principal Investigator, Review of the Recommendations for Delivery of Substance Abuse and Mental Health Services, Curricula, and Training Models for Working in Managed Care and Other Primary Care Settings, Substance Abuse and Mental Health Services Administration/ Center for Substance Abuse Treatment	1997 - 1998

## **Selected Publications**

### ***Peer-reviewed Articles***

- Zarin DA**, Young JL, West JC. Challenges to evidence-based medicine: A comparison of patients and treatments in randomized controlled trials with patients and treatments in a practice research network. *Soc Psychiatry Psychiatr Epidemiol.* 2005 Jan;40(1):27-35.
- Kulasingam SL, Samsa GP, **Zarin DA**, Rutschmann OT, Patwardhan MB, McCrory DC, Schmechel DE, Matchar DB. When should functional neuroimaging techniques be used in the diagnosis and management of Alzheimer's dementia? A decision analysis. *Value Health* 2003 Sept.-Oct.;6(5): 542-50.
- Wang C, Schwaitzberg S, Berliner E, **Zarin DA**, Lau J. Hyperbaric oxygen for treating wounds: a systematic review of the literature. *Arch Surg* 2003; 138:272-279.
- Berliner E, Ozbilgin B, **Zarin DA**. A systematic review of pneumatic compression for treatment of chronic venous insufficiency and venous ulcers. *J Vasc Surg* 2003; 37:539-544.
- Eisenberg JM, Zarin D: Health Technology Assessment in the United States-Past, Present, and Future. *Int J Technology Assess Health Care* 2002; 18:192-198.
- Zarin DA, Pincus HA, West JC, McIntyre JS: Practice-based research in psychiatry. *Am J Psychiatry* 1997; 154: 1199-1208.

### ***Other Publications***

- Zarin, DA.** Using Research to Guide Healthcare Decisions in AAAS Science and Technology Policy Yearbook. Edited by Teich A, Nelson SD, McEnaney C, Lita SJ. American Association for the Advancement of Science, 2001
- Zarin DA**, West JC, Hart C. The American Psychiatric Association's Agenda for Evidence-Based Quality in Improving Mental Health Care Commitment to Quality. Edited by Dickey B and Sederer L. American Psychiatric Publishing, Inc., 2001.
- Zarin DA**, Considerations in Choosing, Using and Interpreting a Measure for a Particular Clinical Context in Handbook of Psychiatric Measures. Edited by Rush JA. Washington DC, American Psychiatric Press 2000.

### ***Practice Guidelines and Technology Assessments***

- Practice Guidelines and Patient Guides on many topics, including: Eating Disorders, Major Depressive Disorder, Bipolar Disorder, Dementia, Anxiety Disorders, Schizophrenia, and Substance Use Disorders.*
- Technology Assessments on variety of topics, including: Fecal Occult Blood Testing, PET scanning for dementia, PET scanning for cancers, Spine Surgery, Vulnerable Plaques, Benign Prostatic Hyperplasia, Knee Arthroscopy, Living Donor Liver Transplantation.*

## ***Annice M Bergeris***

Information Research Specialist

### **Education and Training**

Shepherd College      A.S.      1985      Business

### **Research and Professional Experience**

#### Information Research Specialist

ClinicalTrials.gov web site at the National Library of Medicine. 12/2001 - present

- Liaison for NIH, FDA and Industry pertaining to policy and PRS issues
- Quality Assurance & Compliance for ClinicalTrials.gov data
- Work closely with web developers to design and implement new technical features of the web site
- Analyze and report on ClinicalTrials.gov web site statistics
- Manage the Aspen contract for supporting ClinicalTrials.gov
- Maintain the team's internal web site, including summary reports and contact list

#### Program Analyst

NIH/NIAID/Division of AIDS. 1/1993 – 12/2001

- Design, program and maintain multiple MIS databases for Division of AIDS, merging data from five different Clinical Trials groups with numerous data platforms
- Prepare Annual, Monthly and Weekly Reports for the Division of AIDS and Congress pertaining to demographics of patients, accrual and status of study protocol
- Proficient in SAS, ACCESS, Word, WordPerfect, Excel, Foxpro, Powerpoint, HTML. Design and maintain internal budget spreadsheets for contracts managed by the Division of AIDS
- Design Web pages for multiple branches within the Division, designed and maintained the Adult & Pediatric AIDS Clinical Trials RFA web pages
- Track Site Monitoring Activity between the Site Monitoring group and Division staff
- Responsible for collating answers from Clinical Site Surveys and producing reports on the data
- Prepare plots, graphs and slide for presentation for Advisory Council, Advocacy Groups and AIDS Clinical Trials Group Meetings

#### Application Examiner

FDA/CDER/Div. of Gastrointestinal Drug Products. 8/1987-1/1993

- Prepared necessary correspondence to notify the regulated industry the results of scientific reviews for their submissions
- Prepared complex SQL programs for the Management Information System databases to produce reports for Division staff
- Responded to information requests from the regulated industry and consumers

**Nicholas C. Ide**

Technical Lead, ClinicalTrials.gov

**Education and Training**

Hiram College	B.A.	1983	Computer Science
University of Maryland	M.S.	1987	Computer Science

**Research and Professional Experience**

Chief Architect, Thoughtful Solutions, Inc. (2001-present)

Managed, designed, implemented, and maintained the ClinicalTrials.gov web site for the National Library of Medicine of the National Institutes of Health. Responsibilities included all aspects of the project, including architecture, technology selection, design, implementation, requirements, management of development staff, communications with customer, vendors, and data providers, and maintenance of the production web site. Web site required frequent software updates and daily data updates, while maintaining continuous operation of the site. ClinicalTrials.gov team also developed the Protocol Registration System.

Chief Engineer, Commerce One, Inc. (1985-2001)

Provided programming and financial modeling expertise to Fannie Mae's portfolio management group. Designed and built an award winning book based income model for Fannie Mae portfolio strategy analysts. Identified appropriate requirements and designed and implemented solution in OpenStep. Mentored Fannie Mae technical staff in object oriented programming and Objective C.

Designed and implemented processes and procedures for testing and deploying into operations all of the Business Systems applications as part of the Integration and Test department at INTELSAT. Coordinated with developers, users, and support staff to increase reliability of the operational software environment. Designed, documented, negotiated, marketed and operationalized departmental procedures.

Designed and implemented a database application using Microsoft ACCESS and Visual Basic. Implemented a GUI front-end for an on-line service utilizing Microsoft's Visual C++.

Enhanced AMF AccuScore/Advantage to support forty-frame game. Diagnosed and fixed long standing intermittent problems in communication software and device drivers.

Managed and contributed to the design and implementation of a hand-held consumer electronics product based on the NEC V20. Architecture utilized V20 as master processor with 75304 slave processor for managing external interfaces (keyboard, screen, power, external communications). System was completely testable via automated scripting.

Designed and implemented comprehensive Configuration Management (CM) plans and procedures for all INTELSAT Ground Networks Projects. Task involved researching practices of each project and of INTELSAT Integration and Test group. Designed underlying network to allow centralized storage of files for a heterogeneous network. CM tools were studied, selected, installed, and project staff were trained.

Designed and implemented an Ada interface to the NASA Transportable Application Environment. Resolved language interface issues between C and Ada producing a portable system intuitive for Ada programmers.

Redesigned and implemented the TAE Plus Code Generation facility using C++ to generate code in C, C++, and Ada.

Key technical contributor on development of the INTELSAT Headquarters Subsystem to support Satellite Switched Time Division Multiple Access operations-provided by INTELSAT VI satellite series including requirements analysis through architectural design, detailed design, implementation, and integration.

Responsible for all technical aspects of upgrading a communications satellite operation system from VMS 3.2 to VMS 4.6. Performed analysis of entire INTELSAT IOCTF system and made modifications to application software. Installed products, updated system configuration and command files, and coordinated hardware maintenance. Emphasis was on reliability and minimizing impact on ongoing operations.

Key technical contributor on fixed price enhancements to INTELSAT IOCTF system, which provided for storage and retrieval of realtime data. The data was stored in disk files and could be displayed on graphic display units under operator control.

Ported a number of diagnostic and application programs from ULTRIX to VMS. Programs used the PIXAR SIMD image processing hardware.

Studied a complex portion of Voice of America and produced a detailed document with specific recommendations for computerization.

Ported the NASA/GSFC Transportable Applications Executive from standard BSD 4.2 UNIX to the Cyber VX/VE System V UNIX emulation system.

Enhanced and maintained a database package used with the NASA Landsat Analysis System and the Transportable Applications Executive. Built ANSI standard tape sub-system and ported the entire system to UNIX.

Teaching Assistant, University of Maryland (1984-1985)

Prepared and taught two lectures per week for an introductory computer science course, helped students, and graded papers.

Assistant to the Director of Academic Computing, Hiram College, (1983-1984)

Assisted faculty members in analysis of appropriate CAI applications and worked with programmers on designs and implementations. Provided support to faculty and students with other academic computing needs.

## **Publications**

Gillen JE, Tse T, **Ide NC**, McCray AT. Design, implementation and management of a web-based data entry system for ClinicalTrials.gov. *Medinfo*. 2004;2004:1466-70.

Kayaalp M, Aronson AR, Humphrey SM, **Ide NC**, Tanabe LK. Methods for accurate retrieval of MEDLINE citations in functional genomics. *Proceedings of the Text Retrieval Conference (TREC-2003)*. 2003:175-84.

McCray AT, **Ide NC**, Loane RR, Tse T. Strategies for supporting consumer health information seeking. *Medinfo*. 2004;2004:1152-6.

McCray AT, Dorfman E, Ripple A, **Ide NC**, Jha M, Katz DG, Loane RF, Tse T. Usability issues in developing a Web-based consumer health site. *Proc AMIA Symp* 2000:556-60.

McCray AT, **Ide NC**. Design and implementation of a national clinical trials registry. *J Am Med Inform Assoc*. 2000 May-Jun;7(3):313-23.

Rosemblat G, Tse T, Gemoets D, Gillen JE, **Ide NC**. Supporting access to consumer health information across languages. Submitted to *AMIA 05*.

## ***Tony Tse***

Staff Scientist

### **Education and Training**

Dartmouth College, Hanover, NH	AB	1983-87	Biochemistry
Yale School of Medicine, New Haven, CT		1987-89	
George Washington, Wash DC	MS	1992-96	Information Systems Technology
University of Maryland, College Park	PhD	1996-03	Library & Information Science

### **Research and Professional Experience**

#### ***Employment***

Staff Scientist, Lister Hill National Center for Biomedical Communications, National Library of Medicine	1999-present
Senior Program Analyst, Howard Hughes Medical Institute	1991-97; 98-99
Graduate Research/Teaching Assistant, University of Maryland	1997-98
Biology Science Assistant, National Science Foundation	1990-91

#### ***Honors***

NLM Staff Recognition Awards	2000-03
Eugene Garfield Doctoral Dissertation Fellowship, Beta Phi Mu	2001
ClinicalTrials.gov Team Group Award	2000

### **Recent Publications**

- Divita G, **Tse T**, Roth L. Failure Analysis of MetaMap Transfer (MMTx). Medinfo. 2004;2004:763-7.
- Gemoets D, Rosemblat G, **Tse T**, Logan R. Assessing readability of consumer health information: an exploratory study. Medinfo. 2004:869-73.
- Gillen JE, **Tse T**, Ide NC, McCray AT. Design, Implementation and Management of a Web-based Data Entry System for ClinicalTrials.gov. Medinfo. 2004:1466-70.
- McCray AT, Ide NC, Loane RR, **Tse T**. Strategies for supporting consumer health information seeking. Medinfo. 2004:1152-6.
- McCray AT, **Tse T**. Understanding search failures in consumer health information systems. Proc AMIA Symp 2003: 430-4.
- Rosemblat G, **Tse T**, Gemoets D. Adapting a Monolingual Consumer Health System for Spanish Cross-Language Information Retrieval. Advances in Knowledge Organization, Vol. 9 (2004), p. 315-321; ISKO 2004, London, UK.
- Rosemblat G, Gemoets D, Browne AC, **Tse T**. Machine translation-supported cross language information retrieval for a consumer health resource. Proc AMIA Symp. 2003: 564-8.
- Soergel D, **Tse T**, Slaughter L. Helping healthcare consumers understand: an "interpretive layer" for finding and making sense of medical information. Medinfo. 2004:931-5.
- Tse T**, Soergel D. Exploring medical expressions used by consumers and the media: an emerging view of consumer health vocabularies. Proc AMIA Symp. 2003.
- Tse T**, Soergel D. Procedures for mapping vocabularies from non-professional discourse and a case study from the medical domain. Proc ASIST Annual Meeting. 2003. p. 174-83.