

9-1-2007

# Limitations of Clinical Databases used in Research

Aide Onime

Follow this and additional works at: [https://digitalrepository.unm.edu/hsc\\_ctsc\\_trainee\\_papers](https://digitalrepository.unm.edu/hsc_ctsc_trainee_papers)

---

## Recommended Citation

Onime, Aide. "Limitations of Clinical Databases used in Research." (2007). [https://digitalrepository.unm.edu/hsc\\_ctsc\\_trainee\\_papers/16](https://digitalrepository.unm.edu/hsc_ctsc_trainee_papers/16)

This Article is brought to you for free and open access by the Health Sciences Research Centers at UNM Digital Repository. It has been accepted for inclusion in Clinical and Translational Science Center Trainee Scholarly Output by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

## **Limitations of Clinical Databases used in Research**

**MSCR Biomedical Informatics Level 1 Writing Project, Aide Onime, MD**

### **Introduction**

Clinical research databases are a key foundation to the health sciences. There are important concepts known among information professionals that are not utilized by clinical researchers who build, use or manage these data stores without involving these professionals.(1) This leads to problems in utilization and linking of these databases. This report presents a review of the English literature on database limitations in research.

### **Materials and Methods**

A literature search was conducted in PubMed (National Library of Medicine). The MeSH Database was used to search various combinations of terms. The initial MeSH search term was “Databases”, the subheading “standards” was selected, restricted to major topic heading and limited to last 5 years, humans, English and reviews; 22 articles were found. Next the MeSH terms “Databases”, “Biomedical Research” and “Vocabulary, Controlled” were combined using the “and” function, no restriction to any major topic headings or limits were applied. The result was 17 articles which were downloaded in the MEDLINE text format into a new Endnote library. Other combinations were tried as shown. A leading authority in the controlled vocabulary field of medical informatics was identified “Cimino JJ”. An author search on “Cimino J” was performed in PubMed in Single Citation Matcher and this returned 193 articles, while “Cimino JJ” returned 143 articles. The “OR” function was used to merge both searches in the History mode and this

yielded 193 unique articles. Various limits were then applied as shown. The preferred limits were; published in the last 10 years, items with links to full text, Humans and English, 40 articles met these criteria. Only 10 of these were deemed relevant to the topic. These ten were downloaded into the Endnote library. Full text articles selected from both search strategies were reviewed.

Search	Most Recent Queries	Time	Result
<a href="#">#14</a>	Search "Databases/standards"[Majr] AND "Biomedical Research"[Mesh] Limits: <b>only items with links to full text, Humans, English</b>	1:53:14	<a href="#">14</a>
<a href="#">#13</a>	Search "Databases/standards"[Majr] AND "Biomedical Research"[Mesh] Limits: <b>Humans, English</b>	1:52:50	<a href="#">21</a>
<a href="#">#12</a>	Search "Databases/standards"[Majr] AND "Biomedical Research"[Mesh]	1:47:46	<a href="#">27</a>
<a href="#">#8</a>	Search "Databases"[Mesh] AND "Biomedical Research"[Mesh] AND "Vocabulary, Controlled"[Mesh]	1:38:35	<a href="#">17</a>
<a href="#">#1</a>	Search "Databases/standards"[Majr] Limits: <b>published in the last 5 years, only items with links to full text, Humans, Review, English</b>	1:33:11	<a href="#">22</a>

**Table A: PubMed MeSH Database search**

Search	Most Recent Queries	Time	Result
<a href="#">#10</a>	Select <b>10</b> document(s) Search ((Cimino J[author]) OR (#3)) OR (#1) Limits: <b>published in the last 10 years, only items with links to full text, Humans, English</b>	15:01:10	<a href="#">10</a>
<a href="#">#9</a>		14:55:27	<a href="#">40</a>
<a href="#">#7</a>	Search ((Cimino J[author]) OR (#3)) OR (#1)	14:54:44	<a href="#">194</a>
<a href="#">#5</a>	Select <b>194</b> document(s)	14:43:22	<a href="#">194</a>
<a href="#">#1</a>	Search Cimino J[author]	14:41:47	<a href="#">194</a>
<a href="#">#3</a>	Search Cimino JJ[author]	14:40:02	<a href="#">143</a>

**Table B: PubMed Single Citation Matcher author name search**

## **Discussion**

Access to data is power and the field of biomedical informatics has dramatically increased the types of data available to clinical investigators.(2) To successfully conduct research in this way it is necessary that various databases be linked to the research database system.(3) This may involve linking within hospital systems of clinical and administrative databases or linking to national databases such as Medicare.(2, 3) In considering chronic medical disorders such as rheumatoid arthritis some cite various limitations of randomized controlled trials including short duration, selection of limited number of patients meeting specified criteria, less variability than real world patients and lack of application to certain ethnicities; they propose instead establishing observational prospective databases comprising real world patients.(4) Lastly major contributions have been made to research in the field of human genetics utilizing huge databases, the scale and complexity of which increase in microarray databases.(5, 6)

While databases have become more important to both clinicians and researchers, major issues limiting their use have been raised. These relate to data integrity, security, privacy and usefulness.(1) Communication among databases with the development of standard terminology is a major limitation and a subject of informatics research. Despite known advantages medical terminology models are diverse, expensive to construct and often developed manually by domain experts and linguists; automated high throughput methods are now proposed.(7) The Electronic Data Interchange (EDI) transaction standards of Health Insurance Portability and Accountability Act (HIPAA) seeks to establish common data definitions for use in clinical, administrative and financial data transmission.(8) Examples of HIPAA adopted controlled medical terminologies are:

International Classification of Diseases, 9<sup>th</sup> edition, Clinical Modification (ICD-9 CM), National Drug Codes (NDC), Code on Dental Procedures and Nomenclature (CDT), Health Care Financing Administration Common Procedural Coding system (HCPCS) and Current Procedural Terminology, 4<sup>th</sup> Edition (CPT-4).(8-10) Use of controlled medical terminology in electronic health records in addition to improved efficiency of clinical care provide exciting opportunities in data analysis for retrospective and prospective clinical research and easier knowledge dissemination.(9) Experts in controlled medical terminology designate Logical Observation Identifiers Names and Codes (LOINC) and Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT) as potential sources of clinical code sets; both are useful in encoding a large scope of clinical data to facilitate data integration and analysis.(8) A weakness of data entry at the point of care using controlled medical terminologies is the limited inter-coder agreement.(10) Privacy and security concerns including informed consent and confidentiality also arise in the utilization of clinical and research databases. Ethical questions can be especially complex in tissue repositories and genetic databases, proposed solutions include data de-identification, institutional review board (IRB) approval of projects, patient controlled electronic medical records, patient control over disclosure process, input from policymakers and legislators and researcher responsibility and education.(11, 12)

### **Conclusion**

There are important problems particularly with communication and standards which are encountered in the linkage of various administrative, clinical and research databases. This is an ongoing field of important biomedical informatics research.

## References

1. Waldren SE, Quetsch J. Clinical research databases: getting from current state to best practice. *AMIA Annu Symp Proc.* 2003:1043.
2. Neugut AI. Hospital information systems and clinical cancer research. *Cancer Invest.* 2004;22(2):326-7.
3. Ackerman IN, Osborne RH. Integrating data to facilitate clinical research: a case study. *Inform Prim Care.* 2005;13(4):263-70.
4. Yamanaka H, Tohma S. Potential impact of observational cohort studies in Japan on rheumatoid arthritis research and practice. *Mod Rheumatol.* 2006;16(2):75-6.
5. Lowrance WW. The promise of human genetic databases. *BMJ.* 2001 Apr 28;322(7293):1009-10.
6. Stoeckert CJ, Jr., Causton HC, Ball CA. Microarray databases: standards and ontologies. *Nat Genet.* 2002 Dec;32 Suppl:469-73.
7. Zhou L, Tao Y, Cimino JJ, Chen ES, Liu H, Lussier YA, et al. Terminology model discovery using natural language processing and visualization techniques. *J Biomed Inform.* 2006 Dec;39(6):626-36.
8. Dykes PC, Currie LM, Cimino JJ. Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway. *J Biomed Inform.* 2003 Aug-Oct;36(4-5):313-25.
9. Chiang MF, Casper DS, Cimino JJ, Starren J. Representation of ophthalmology concepts by electronic systems: adequacy of controlled medical terminologies. *Ophthalmology.* 2005 Feb;112(2):175-83.
10. Hwang JC, Yu AC, Casper DS, Starren J, Cimino JJ, Chiang MF. Representation of ophthalmology concepts by electronic systems: intercoder agreement among physicians using controlled terminologies. *Ophthalmology.* 2006 Apr;113(4):511-9.
11. Auray-Blais C, Patenaude J. A biobank management model applicable to biomedical research. *BMC Med Ethics.* 2006;7:E4.
12. Kohane IS, Altman RB. Health-information altruists--a potentially critical resource. *N Engl J Med.* 2005 Nov 10;353(19):2074-7.