Building the Informatics Infrastructure for Comparative Effectiveness Research (CER)

A Review of the Literature

Marianne Hamilton Lopez, MPA,* Erin Holve, PhD, MPH, MPP,*
Indra Neil Sarkar, PhD, MLIS,† and Courtney Segal, BA*

Background: Technological advances in clinical informatics have made large amounts of data accessible and potentially useful for research. As a result, a burgeoning literature addresses efforts to bridge the fields of health services research and biomedical informatics. The Electronic Data Methods Forum review examines peer-reviewed literature at the intersection of comparative effectiveness research and clinical informatics. The authors are specifically interested in characterizing this literature and identifying cross-cutting themes and gaps in the literature.

Methods: A 3-step systematic literature search was conducted, including a structured search of PubMed, manual reviews of articles from selected publication lists, and manual reviews of research activities based on prospective electronic clinical data. Two thousand four hundred thirty-five citations were identified as potentially relevant. Ultimately, a full-text review was performed for 147 peer-reviewed papers.

Results: One hundred thirty-two articles were selected for inclusion in the review. Of these, 88 articles are the focus of the discussion in this paper. Three types of articles were identified, including papers that: (1) provide historical *context* or frameworks for using clinical informatics for research, (2) *describe platforms and projects*, and (3) discuss issues, challenges, and applications of *natural language processing*. In addition, 2 cross-cutting themes emerged: the challenges of conducting research in the absence of standardized ontologies and data collection; and unique data governance concerns related to the transfer, storage, deidentification, and access to electronic clinical data. Finally, the authors identified several current gaps on important topics such as the use of clinical informatics for cohort identification, cloud computing, and single point access to research data.

Key Words: comparative effectiveness research, clinical informatics, electronic clinical data, electronic health records, data sharing

(*Med Care* 2012;50: S38–S48)

Technological advances in clinical informatics have made large amounts of data accessible and potentially useful for research. As a result, a burgeoning literature addresses efforts to bridge the fields of health services research (HSR) and biomedical informatics. The goal of the Academy Health Electronic Data Methods (EDM) Forum is to facilitate learning and foster collaboration by working with 11 comparative effectiveness research (CER) projects [supported through 3 grant programs: Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies (PROSPECT), Scalable Distributed Research Network (DRN) for CER, and the Enhanced Registry for Quality Improvement (QI) and CER] designed to build infrastructure and conduct CER with electronic clinical data (ECD). CER aims to determine what works best for whom and under what conditions[1] and observational CER studies may leverage large amounts of data from numerous systems and sources. Clinical informatics platforms, models, and tools are being developed to support new infrastructure and studies across multiple sites and sources. As these efforts develop, it is important to understand the current state of the peer-reviewed literature at the intersection of CER and clinical informatics. The EDM Forum literature review aims to characterize this new body of literature on CER and clinical informatics, as well as identify cross-cutting themes and gaps in the literature.

METHODS

The relatively new use of the term CER and the number of concepts related to CER can complicate efforts to develop an automated search for relevant articles. The National Library of Medicine has developed an initial strategy to identify CER in the peer-reviewed literature.[2] According to PubMed search strategies applied in 2010, literature annotated with the Medical Subject Headings (MeSH) term for "comparative effectiveness research" as a topic generates extremely small results (n = 375), whereas a search string classifying CER concepts tends to generate extremely large results (2 million citations or more).[3] As a result, before this review, no existing PubMed[4] search strategy has identified relevant articles on CER and clinical informatics (defined here as a specific subdiscipline of biomedical informatics that focuses on informatics in the clinical context as opposed to molecular biology or public health). On the basis of the input

from librarians at the National Library of Medicine and the University of Vermont, we developed a multistep approach to search and identify relevant articles.

Study Design

A 3-step, curated approach was used to identify relevant articles (the numbers of records retrieved per search are in parentheses), including:

(1) Searching PubMed using MeSH terms (n = 68). This effort focused on a set of search strings, MeSH terminology (including major headings "MH" and minor headings "mh"), and terms known to be associated with key projects, programs, and authors. In addition, using Boolean search techniques (AND, OR, NOT) enabled us to narrow the search. For example, one specific PubMed search used the following search string: (("Informatics" [tiab] OR "Informatics" [MH] OR "Medical Informatics" [MH]) OR ("data mining" [MH] OR "information storage and retrieval" [mh])) AND ("Comparative Effectiveness" [tiab] OR "Comparative Effectiveness Research" [MH]). An additional keyword (KW) search was conducted for the "Learning Healthcare System," generating 7 results.

(2) Manual review of papers from select publication lists was then conducted, including articles referenced in: the PROSPECT, DRN, and Enhanced Registries study proposals (~1500 citations); an annotated bibliography developed for the AcademyHealth project, Health IT for Actionable Knowledge (n = 40)[5]; and a subset of papers presented at the 2010 American Medical Informatics Association Symposium (n = 2).

(3) The third step focused on reviewing articles identified in publication lists and/or Web sites from research activities related to the EDM Forum (ie, research based on prospective ECD). This list of projects was developed based on input and discussions with experts working on CER and informatics, including members of the EDM Forum Steering Committee. Projects include: caBIG (cancer Biomedical Informatics Grid),[6] DARTNet,[7] DEcIDE (Developing Evidence to Inform Decisions about Effectiveness),[8] HMORN (HMO Research Network),[9] iDASH (integrating data for analysis, anonymization, and sharing),[10] i2b2 (Informatics for Integrating Biology and the Bedside),[11] OMOP (Observational Medical Outcomes Partnership),[12] PhysioMIMI (Multi-Modality, Multi-Resource Information Integration environment),[13] REDCap (Research Electronic Data Capture),[14] Sentinel Initiative and Mini-Sentinel,[15] SHARP Program (Strategic Health IT Advanced Research Projects),[16] TRIAD (OSU Clinical and Translational Science Awards),[17] and VINCI (VA Informatics and Computing Infrastructure).[18] Review of the available publication lists and project Web sites resulted in 818 potentially relevant articles. Although large datasets are an important source of information on health outcomes, the objective of the review is identifying activities focused on clinical informatics using ECD for research. As a result, this list of projects does not include research on large retrospective datasets that have been mined to answer effectiveness questions (eg, Medicare data files, surgical QI project data files, etc.).

Finally, to test our ability to identify key concepts that did not appear frequently in our searches, we conducted sensitivity tests of possible gaps by conducting individual PubMed searches, combining keyword and MeSH searches for specific key terms that did not appear to have been addressed.

Review and Coding

A total of 2435 citations were potentially relevant for the review. The initial article set (n = 2435) was reduced (to ~400) based on a manual review of titles and abstracts by a formally trained biomedical informatician (I.N.S.). The goal of this initial filtering process was to identify an unduplicated set of articles focused on the set of key concepts developed by the authors that reflect major topics at the intersection of clinical informatics and CER (Table 1).

To select the most relevant subset of articles, articles were excluded if they were not explicitly related to clinical informatics, or focused on genomic rather than clinical data, or clinical outcomes rather than the use of clinical informatics for research. The set of 400 citations then underwent another review of titles and abstracts using the exclusion criteria articulated above resulting in 147 articles. A full-text review was performed for 147 articles and the previously described exclusion criterion was used, resulting in 132 articles selected for discussion.

This final set of 132 articles was divided between 2 reviewers who abstracted and coded their assigned papers, using a standardized abstraction form designed specifically for the EDM Forum literature review. Because the goal of conducting CER with ECD brings together several multidisciplinary fields—specifically, bioinformatics and HSR—it was important to develop a list of salient concepts that are cross-cutting. The list of concepts was developed by a multidisciplinary team comprised of a biomedical informatician (I.N.S.) and HSR researcher (E.H.). The list was then reviewed by EDM Forum consultants, including a clinician and statistician; and members of the EDM Forum Steering Committee, many of whom are engaged in developing infrastructure and methods for conducting CER with ECD.

The list of key concepts was then used to code the review findings and characterize areas of focus, or gaps in the current literature. After the records were retrieved, each

TABLE 1. List of Concepts Used to Code the Primary Folder of Articles on Clinical Informatics and CER

Context	Metadata
Platforms, projects	Patient involvement
Clinical informatics platforms	IRBs
Clinical informatics projects	Governance
Natural language processing	Library of phenotypes
Data use and quality	The Learning Healthcare System and CER
Research networks	
Standardized data collection	Single point access
Identifiers and deidentification	Cloud computing
Security	Cohort identification

CER indicates comparative effectiveness research.

paper was coded with a primary area of focus to quantitatively assess the extent to which specific topics are currently addressed in the literature. Throughout the coding process, inconsistencies between the reviewers were deliberated and the list of concepts was refined. The final list of concepts is shown in Table 1. To validate the coding process, the reviewers randomly selected one third of the articles that were initially reviewed by the other reviewer, and conducted a second, independent review to compare categorization, resulting in a 77% rate of agreement and (Cohen's) κ of 0.78, between the 2 reviewers.

RESULTS

One hundred thirty-two articles were selected for inclusion in the review (see Appendix for a list of selected references). Of these, 88 articles are the focus of the discussion in this paper.

Three *types* of articles were identified as major areas of focus for the current body of literature, including papers that: (1) provide historical *context* or frameworks for using clinical informatics for research, (2) *describe platforms and projects*, and (3) discuss issues, challenges, and applications of *natural language processing* (NLP). In addition, 2 cross-cutting themes emerged: the challenges of conducting research in the absence of standardized ontologies and data collection; and unique data governance concerns related to the transfer, storage, deidentification, and access to ECD. Finally, the authors identified current gaps on topics such as the use of clinical informatics for cohort identification, cloud computing, and single point access to research data.

Characterizing the Literature

Historical Context and Frameworks

Twenty-five papers (19%) provided context for the evolution of clinical informatics used to conduct CER. Taken as a whole, the papers present an overview of the history, challenges, and potential for conducting CER with ECD.

Three articles provide historical insights into early intersections between HSR and biomedical informatics, and the development of clinical informatics tools to conduct CER. The articles include foundational works on CER[19] and papers defining keywords and scope relevant to using biomedical informatics for research.[20,21]

One emerging theme within this category is that multidisciplinary and multisite collaborations can enable new partnerships, which may in turn expand the availability of data, strengthen project sustainability, lead to research innovation, and minimize duplication of effort. Twelve of the articles address the opportunities and challenges of fostering and conducting multidisciplinary and multisite collaborations in HSR and biomedical informatics,[22–35] with 2 articles specifically calling for cross-training between the fields.[22,35]

In one example, McCray et al's[23] review of several International Medical Informatics Association (IMIA) working group meetings illustrates the importance of multidisciplinary collaborations. The 1988 meeting included representation from linguistics, natural language analysis, medical decision-making, knowledge representation, and computer science, demonstrating the multidisciplinary nature of perspectives that influence the area of study. The authors' summary of papers produced from the IMIA's 1994 meeting reflects the value of convening researchers from diverse disciplines in the effort to "model[ing] medical knowledge" for a variety of uses in clinical care and research. The authors further note the collaborative nature of researchers in this arena, commenting on IMIA participants' interest in discussing "the importance of sharing knowledge and tools and ... continued and increased efforts to develop sharable research products."

Yet, multidisciplinary and multisite collaborations can give rise to new challenges such as working with and obtaining approval from multiple IRBs.[9] One perspective that emerged in the papers addressing multidisciplinary and multisite collaborations is that current funding and publishing practices offer few incentives to share data, which could negatively influence researchers' readiness to engage in collaborative efforts.[24,26] According to de Carvalho et al[24] resistance to data sharing could negatively impact research:

In a system that emphasizes competition rather than collaboration among researchers, data sets resulting from multimillion investments from tax payers sit idle inside locked computers, only available to a small number of researchers despite their containing the seeds that would allow for the exploration of a vast number of important research questions that could change the healthcare landscape.

Platforms and Projects

Thirty papers (23%) describe the aims, content, functionality, and methods used by a variety of ongoing clinical informatics projects and platforms. Principally, the literature addressed the most well-established efforts, including: caBIG,[36–39] DARTNet,[40] HMORN,[41–44] i2b2,[31,45–47] OMOP,[42] REDCap,[48] Sentinel Initiative and Mini-Sentinel,[42,49,50] and TRIAD.[51] These papers provide an overview of various efforts that are underway using clinical informatics for research, many of which include information on the projects' areas of focus; provide progress updates; or describe tools and platforms that have been developed or are in development.

NLP

NLP is used to extract relevant data from the free text embedded in electronic health records and text documents and is used in a variety of contexts, most commonly in HSR to conduct drug and vaccine surveillance, or effectiveness research. Thirty articles (23%) focus on NLP tools.

The literature on NLP focuses on tools for extracting relevant data from a variety of text sources and generating deidentified data applicable to the needs of research.[52] Articles also describe the range of NLP techniques and frameworks being developed, used, modified, and evaluated.[47,53–78]

Ten papers discuss the i2b2 NLP data mining challenges, in which a variety of NLP techniques (eg, machine learning and rule-based approaches) were tested using the same set of data in an effort to develop new approaches for extracting information from clinical text. Three papers discuss the 2006 Smoking Challenge[29,30,79] and 7 discuss the 2008 Obesity

Challenge.[31–36,80] In his overview of the Obesity Challenge, Uzuner[80] describes that the challengers used NLP using 1237 discharge summaries annotated by 2 obesity experts. The annotations consisted of textual judgments (the classification of each disease based on explicit information in the documents) or intuitive judgments (the classification of each disease based on the expert's intuition and judgment), and were evaluated using microaveraged and macroaveraged precision, recall, and *F*-measure (a statistical measure that considers both precision and recall). Rule-based approaches contributed to top performance within the textual category, whereas the machine learning approaches were more successful within the intuitive test.

Emerging Themes
The Need for Standardization

Conducting CER with ECD may involve data from different clinical informatics platforms and sources, and therefore, diverse systems with different ontologies[81] and varying data collection practices must be anticipated. The need to standardize ontologies and data collection so that researchers can work together in multisite, multidisciplinary research collaborations is a prevalent theme.

Eight papers[28,37,82–87] focus on the need for and/or application of standardized ontologies. Of these, Chute[83] highlights the challenges of establishing common vocabularies because of the increasing number of naming mechanisms and resulting proliferation of medical terms in the field. The flexibility of the number of available clinical terminology systems coupled with the lack of standardized vocabulary for medical terms, has resulted in multiple names for single concepts. The author notes that this issue complicates researchers' ability to share health information across sites.

The need to standardize data collection is an additional issue raised by the authors. Differences in ontologies, informatics platforms, and data entry practices contribute to the complexity of collecting and analyzing multisite electronic data for research. Three papers focus on the need to standardize data collection by applying common data elements.[88–90] Seven papers[27,40,49,91–94] highlight the challenges associated with collecting and analyzing data from multiple sites and informatics platforms, such as the differences in the data collected for clinical use versus research use,[27] variances between the functionality of different EHRs,[91] lack of standardization in how data are reported among multiple data sources,[40,49,92,93] and difficulty in implementing standardization efforts after systems are operational.[94]

Data Governance and Access

CER based on ECD presents unique data governance concerns related to the transfer and storage,[43,92,93,95–97] deidentification,[98–102] and access of ECD.[103–105] At present, the literature focuses on 2 primary issues: the need to ensure the security of transfer and storage of data; and the tension between preserving privacy (often by deidentifying research data) and enabling research data access.

In the literature to date, no consensus has emerged with respect to best practices for data security, especially regarding DRN or centralized data models.[106] Proponents of DRNs argue that the model allows for local control because data never leave their home institution.[43,95–97] In contrast, proponents of centralized data models assert that a centralized data warehouse can be more secure.[92,93]

Authors have also debated whether existing privacy regulations (ie, the Health Insurance Portability and Accountability Act, or HIPAA) provide sufficient protection to patients while balancing access to data for researchers. Two articles focus on the current standards for the protection and use of personal health information, arguing that research with deidentified data poses minimal risk.[103,104] Five of the articles addressed the vulnerabilities of current deidentification efforts.[98–102] Informed consent is also discussed as an issue impacting the availability of ECD for research. Goldstein[105] discusses ways informed consent may be adapted in a changing climate of health care delivery, particularly with the ability for EHRs to strengthen security protocols and ensure patients are informed of various uses of their data.

Identification of Gaps

As discussed, the set of concepts initially developed by the authors reflect key topics at the intersection of clinical informatics and CER (Table 1). To the extent that the reviewers did not identify *any* relevant articles that aligned with the codes/concepts developed for the review (even after conducting specialized keyword searches), the authors assessed these concepts as being "gaps' in the literature. No articles were identified as focusing on "cohort identification," "cloud computing," or "single point access" to research data. In some instances, each of these concepts were discussed as a challenge or issue in the context of papers on other topics, but were not the primary focus of the article.

DISCUSSION

The literature review demonstrates the degree to which the disciplines and fields focused on CER and clinical informatics are discussing approaches to advance the use of clinical informatics for research to improve patient outcomes. For example, the literature review illustrates the variety of approaches that use different data models or network strategies for CER. The range of efforts exemplifies the current focus on exploration and innovation in this arena. Similarly, differences of opinion regarding the utility of deidentifying data for research, and the associated privacy risks that may emerge, come to the fore. All of these crosscutting themes reflect a nascent, but rich, discussion, and the breadth of perspectives in this growing community of scientists engaged in expanding the current paradigm of effectiveness research. The presence of notable gaps in the literature (eg, informatics strategies for cohort identification) underscores new opportunities for scholarship.

This first effort to characterize the literature relevant for CER and clinical informatics clearly demonstrates that there is no one discipline or field whose taxonomy can be used to readily access the relevant literature. For the near future, at least, the need to review important issues, tools,

and techniques, and cross-reference key investigators is necessary to assemble the literature that is relevant to building the CER infrastructure. However, some limitations are inherent in this effort. For example, some topics identified as gaps in the literature may be because of limitations in our current search or coding processes. Another potential limitation is that the literature on clinical informatics and CER is still at its infancy and that some concepts of interest to CER investigators, while innovative and important, are not well established or may not have achieved peer-reviewed status. It is possible that some innovative approaches may be identified in the grey literature (research papers and reports that are not peer reviewed[107]). As a result, the EDM Forum is in the process of conducting a structured review of the grey literature to identify additional resources.

The EDM Forum has also made strides to develop and disseminate resources that facilitate ongoing review of the literature, including a set of PubMed search terms that provide a starting point for exploration; the EDM Forum's abstraction form, which was created for this review; and an annotated bibliography of all the selected articles. A glossary of relevant terms has also been developed as part of www.edm-forum.org. Through these and related efforts we will continue to monitor the evolution of the literature on CER and clinical informatics.

Over the next 2 years, an influx of new research supported by the American Recovery and Reinvestment Act (ARRA) of 2009 and the Health Information Technology for Economic and Clinical Health (HITECH) Act is likely to result in new scholarship, particularly with respect to CER studies using ECD. Charting the progress of this emerging scientific endeavor promises unique and interesting challenges, and new opportunities for discovery that may only be achieved by bringing diverse literatures together. To this end, EDM Forum staff will be conducting an ongoing scan of the literature, in order to gather the latest published data on the intersection of clinical informatics and CER.

As part of this effort the authors are collaborating to develop an automated approach (ie, identification of keyword or MeSH descriptors that can be used with high reliability to identify relevant literature) that can facilitate our ability to identify relevant articles with greater precision. Of particular interest is developing methods to detect literature associated with gaps in the review.

CONCLUSIONS

Identifying literature that is relevant to clinical informatics and CER requires a clear conceptual framework of the particular issues that are most relevant to explore. The present review of CER and clinical informatics is unique in its effort to bring together literature from a number of domains, including technology, medicine, research, and policy.

Having characterized this initial literature, including multidisciplinary viewpoints and ongoing initiatives dedicated to developing an infrastructure for collecting and analyzing prospective ECD, the authors hope these materials will be useful to other investigators engaged in new approaches to conduct CER with ECD. Ideally, this review and the underlying approach will serve as the foundation for researchers to access relevant literature on clinical informatics for CER. This initial review is a first step toward aggregating and characterizing scholarship in CER and clinical informatics, as well as identifying themes and gaps that may be critical to address as part of an overarching strategy aimed at advancing science to improve patient outcomes.

ACKNOWLEDGMENTS

Appendix. Summary Table of Articles on CER and Clinical Informatics Selected for Review and Discussion

Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. *J Clin Oncol.* 2010;28:4268–4274.

Abernethy AP, Wheeler JL, Zafar SY. Management of gastrointestinal symptoms in advanced cancer patients: the rapid learning cancer clinic model. *Curr Opin Support Palliat Care.* 2010;4:36–45.

Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, et al. Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. *Med Care.* 2009;47(10):1091–1097.

Ambert KH, Cohen AM. A system for classifying disease comorbidity status from medical discharge summaries using automated hotspot and negated concept detection. *J Am Med Inform Assoc.* 2009;16:590–595.

Amin W, Singh H, Pople AK, et al. A decade of experience in the development and implementation of tissue banking informatics tools for intra and inter-institutional translational research. *J Pathol Inform.* 2010;1:12. Available at: http://www.jpathinformatics.org/temp/JPatholInform1112-3704914_101729.pdf.

Aronsky D, Haug PJ. Assessing the quality of clinical data in a computer-based record for calculating the pneumonia severity index. *J Am Med Inform Assoc.* 2000;7:55–65.

Atkins D. Connecting research and patient care: lessons from the VA's Quality Enhancement Research Initiative. *J Gen Intern Med.* 2010;25:1–2.

Bates DW, Pappius E, Kuperman GJ, et al. Using information systems to measure and improve quality. *Int J Med Inform.* 1999;53:115–124.

Behrman RE, Benner JS, Brown JS, et al. Developing the Sentinel System—a national resource for evidence development. *N Engl J Med.* 2011;364:498–499.

Bellin E, Fletcher DD, Geberer N, et al. Democratizing information creation from health care data for quality improvement, research, and education—the Montefiore Medical Center Experience. *Acad Med.* 2010;85:1362–1368.

Berger ML, Mamdani M, Atkins D, et al. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I. *Value Health.* 2009;12:1044–1052.

Black N. Secondary use of personal data for health and health services research: why identifiable data are essential. *J Health Serv Res Policy.* 2003;8:36–40.

Bramsen P, Deshpande P, Lee YK, et al. Finding temporal order in discharge summaries. *AMIA Annu Symp Proc.* 2006:81–85.

Brennan PF, Stead WW. Assessing data quality: from concordance, through correctness and completeness, to valid manipulatable representations. *J Am Med Inform Assoc.* 2000;7:106–107.

Brown JS, Holmes JH, Shah K, et al. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care.* 2010;48:S45–S51.

Brown JS, Moore KM, Braun MM, et al. Active influenza vaccine safety surveillance: potential within a healthcare claims environment. *Med Care.* 2009;47:1251–1257.

Bu D, Pan E, Walker J, et al. Benefits of information technology-enabled diabetes management. *Diabetes Care.* 2007;30:1137–1442.

Buetow KH. An infrastructure for interconnecting research institutions. *Drug Discov Today.* 2009;14:605–610.

Carpenter PC, Chute CG. The Universal Patient Identifier: a discussion and proposal. *Proc Annu Symp Comput Appl Med Care.* 1993:49–53.

Carrell D, Miglioretti D, Smith-Bindman R. Coding free text radiology reports using the Cancer Text Information Extraction System (caTIES). *AMIA Annu Symp Proc.* 2007;2007:889.

de Carvalho ECA, Batilana AP, Simkins J, et al. Application description and policy model in collaborative environment for sharing of information on epidemiological and clinical research data sets. *PLoS One.* 2010;5:e9314.

Chan KA. Development of a multipurpose dataset to evaluate potential medication errors in ambulatory settings. In: Henriksen K, Battles JB, Marks ES, et al. eds. Advances in Patient Safety: From Research to Implementation (Vol 2: Concepts and Methodology). Rockville, MD: Agency for Healthcare Research and Quality; 2005:225–238.

Chen RT, Glasser JW, Rhodes PH, et al. Vaccine Safety Datalink project: a new tool for improving vaccine safety monitoring in the United States. The Vaccine Safety Datalink Team. *Pediatrics.* 1997;99:765–773.

Chute CG, Beck SA, Fisk TB, et al. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *J Am Med Inform Assoc.* 2010;17:131–135.

Chute CG, Elkin PL, Sherertz DD, et al. Desiderata for a clinical terminology server. *Proc AMIA Symp.* 1999:42–46.

Chute CG, Yang Y, Evans DA. Latent Semantic Indexing of medical diagnoses using UMLS semantic structures. *Proc Annu Symp Comput Appl Med Care.* 1991:185–189.

Chute CG. Clinical classification and terminology: some history and current observations. *J Am Med Inform Assoc.* 2000;7:298–303.

Chute CG. The horizontal and vertical nature of patient phenotype retrieval: new directions for clinical text processing. *Proc AMIA Symp.* 2002:165–169.

Cimino JJ. Terminology tools: state of the art and practical lessons. *Methods Inf Med.* 2001;40:298–306.

Clark C, Good K, Jezierny L, et al. Identifying smokers with a medical extraction system. *J Am Med Inform Assoc.* 2008;15:36–39.

Corn M, Rudzinski KA, Cahn MA. Bridging the gap in medical informatics and health services research: workshop results and next steps. *J Am Med Inform Assoc.* 2002;9:140–143.

Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value Health.* 2009;12:1053–1061.

D'Avolio LW, Bui AA. The Clinical Outcomes Assessment Toolkit: a framework to support automated clinical records-based outcomes assessment and performance measurement research. *J Am Med Inform Assoc.* 2008;15:333–340.

D'Avolio LW, Farwell WR, Fiore LD. Comparative effectiveness research and medical informatics. *Am J Med.* 2010;123:e32–e37.

D'Avolio LW. Electronic medical records at a crossroads: impetus for change or missed opportunity? *JAMA.* 2009;302:1109–1111.

de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Fam Pract.* 2006;23:253–263.

Deitzer JR, Payne PR, Starren JB. Coverage of clinical trials in existing ontologies. *AMIA Annu Symp Proc.* 2006:903.

Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform.* 2009;42:760–772.

Deshmukh VG, Meystre SM, Mitchell JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. *BMC Med Res Methodol.* 2009;9:70. Available at: http://www.biomedcentral.com/content/pdf/1471-2288-9-70.pdf.

Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. *Health Aff (Millwood).* 2009;28:454–466.

Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and ethical considerations for linking clinical and administrative databases. *Am Heart J.* 2009;157:971–982.

Eisen S, Francis J. Transformation of VHA health data into clinically useful information to provide quality veteran care. *J Rehabil Res Dev.* 2010;47:xiii–xv.

Appendix. Summary Table of Articles on CER and Clinical Informatics Selected for Review and Discussion (*continued*)

El Emam K, Jabbouri S, Sams S, et al. Evaluating common de-identification heuristics for personal health information. *J Med Internet Res.* 2006;8:e28.

El-Ghatta SB, Cladé T, Snyder JC. Integrating clinical trial imaging data resources using service-oriented architecture and grid computing. *Neuroinformatics.* 2010;8:251–259.

Elkin PL, Mohr DN, Tuttle MS, et al. Standardized problem list generation, utilizing the Mayo canonical vocabulary embedded within the Unified Medical Language System. *Proc AMIA Annu Fall Symp.* 1997:500–504.

Elkins JS, Friedman C, Boden-Albala B, et al. Coding neuroradiology reports for the Northern Manhattan Stroke Study: a comparison of natural language processing and manual review. *Comput Biomed Res.* 2000;33:1–10.

Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. *J Am Med Inform Assoc.* 2009;16:316–327.

Etheredge LM. Creating a high-performance system for comparative effectiveness research. *Health Aff (Millwood).* 2010;29:1761–1767.

Farkas R, Szarvas G, Hegedus I, et al. Semi-automated construction of decision rules to predict morbidities from clinical texts. *J Am Med Inform Assoc.* 2009;16:601–605.

Fefferman NH, O'Neil EA, Naumova EN. Confidentiality and confidence: is data aggregation a means to achieve both?. *J Public Health Policy.* 2005;26:430–449.

Fridsma DB, Evans J, Hastak S, et al. The BRIDG project: a technical report. *J Am Med Inform Assoc.* 2008;15:130–137.

Friedlin J, Grannis S, Overhage JM. Using natural language processing to improve accuracy of automated notifiable disease reporting. *AMIA Annu Symp Proc.* 2008:207–211.

Friedman C, Shagina L, Lussier Y, et al. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc.* 2004;11:392–402.

Goldstein MM. Health information technology and the idea of informed consent. *J Law Med Ethics.* 2010;38:27–35.

Goryachev S, Kim H, Zeng-Treitler Q. Identification and extraction of family history information from clinical reports. *AMIA Annu Symp Proc.* 2008:247–251.

Grant A, Moshyk A, Diab H, et al. Integrating feedback from a clinical data warehouse into practice organisation. *Int J Med Inform.* 2006;75:232–239.

Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42:377–381.

Hastings S, Oster S, Langella S, et al. Adoption and adaptation of caGrid for CTSA. *Summit on Translat Bioinforma.* 2009;2009:44–48.

Hazlehurst B, Frost HR, Sittig DF, et al. MediClass: a system for detecting and classifying encounter-based clinical events in any electronic medical record. *J Am Med Inform Assoc.* 2005;12:517–529.

Hazlehurst B, Naleway A, Mullooly J. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. *Vaccine.* 2009;27:2077–2083.

Hazlehurst B, Sittig DF, Stevens VJ, et al. Natural language processing in the electronic medical record: assessing clinician adherence to tobacco treatment guidelines. *Am J Prev Med.* 2005;29:434–439.

Himes BE, Dai Y, Kohane IS, et al. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *J Am Med Inform Assoc.* 2009;16:371–379.

Honigman B, Lee J, Rothschild J, et al. Using computerized data to identify adverse drug events in outpatients. *J Am Med Inform Assoc.* 2001;8:254–266.

Hynes DM, Perrin RA, Rappaport S, et al. Informatics resources to support health care quality improvement in the veterans health administration. *J Am Med Inform Assoc.* 2004;11:344–350.

Jiang G, Solbrig HR, Iberson-Hurst D, et al. A collaborative framework for representation and harmonization of clinical study data elements using Semantic MediaWiki. *AMIA Summits Transl Sci Proc.* 2010;2010:11–15.

Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III. *Value Health.* 2009;12:1062–1073.

Krishna R, Kelleher K, Stahlberg E. Patient confidentiality in the research use of clinical medical databases. *Am J Public Health.* 2007;97:654–658.

Kukafka R, Ancker JS, Chan C, et al. Redesigning electronic health record systems to support public health. *J Biomed Inform.* 2007;40:398–409.

Kush RD, Helton E, Rockhold FW, et al. Electronic health records, medical research, and the Tower of Babel. *N Engl J Med.* 2008;358:1738–1740.

Langella S, Hastings S, Oster S, et al. Sharing data and analytical resources securely in a biomedical research grid environment. *J Am Med Inform Assoc.* 2008;15:363–373

Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. *BMC Public Health.* 2006;6:235. Available at: http://www.biomedcentral.com/content/pdf/1471-2458-6-235.pdf.

Liao KP, Cai T, Gainer V, et al. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken).* 2010;62:1120–1127.

Los RK, van Ginneken AM, van der Lei J. OpenSDE: a strategy for expressive and flexible structured data entry. *Int J Med Inform.* 2005;74:481–490.

Mandl KD, Lee TH. Integrating medical informatics and health services research: the need for dual training at the clinical health systems and policy levels. *J Am Med Inform Assoc.* 2002;9:127–132.

Manion FJ, Robbins RJ, Weems WA, et al. Security and privacy requirements for a multi-institutional cancer research data grid: an interview-based study. *BMC Med Inform Decis Mak.* 2009;9:31. Retrieved from http://www.biomedcentral.com/content/pdf/1472-6947-9-31.pdf.

Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. *Ann Intern Med.* 2009;151:341–344.

McConnell P, Dash RC, Chilukuri R, et al. The cancer translational research informatics platform. *BMC Med Inform Decis Mak.* 2008;8:60.

McCormick PJ, Elhadad N, Stetson PD. Use of semantic features to classify patient smoking status. *AMIA Annu Symp Proc.* 2008:450–454.

McCray AT, Scherrer JR, Safran C, et al. Concepts, knowledge, and language in health-care information systems. *Methods Inf Med.* 1995;34:1–4.

McDonald CJ, Dexter P, Schadow G, et al. SPIN query tools for de-identified research on a humongous database. *AMIA Annu Symp Proc.* 2005:515–519.

McDonald CJ, Overhage JM, Tierney WM, et al. The Regenstrief Medical Record System: a quarter century experience. *Int J Med Inform.* 1999;54:225–253.

McGraw D, Dempsey JX, Harris L, et al. Privacy as an enabler, not an impediment: building trust into health information exchange. *Health Aff (Millwood).* 2009;28:416–427.

McKinney M. "Huge potential" for EHRs and comparative effectiveness. *Hosp Health Netw.* 2010;84:41–42.

(*Continued*)

Appendix. Summary Table of Articles on CER and Clinical Informatics Selected for Review and Discussion (*continued*)

McMurry AJ, Gilbert CA, Reis BY, et al. A self-scaling, distributed information architecture for public health, research, and clinical care. *J Am Med Inform Assoc.* 2007;14:527–533.

Meystre SM, Friedlin FJ, South BR, et al. Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC Med Res Methodol.* 2010;10:70. Retrieved from http://www.biomedcentral.com/content/pdf/1472-6947-8-60.pdf.

Mishra NK, Cummo DM, Arnzen JJ, et al. A rule-based approach for identifying obesity and its comorbidities in medical discharge summaries. *J Am Med Inform Assoc.* 2009;16:576–579.

Mohanty SK, Mistry AT, Amin W, et al. The development and deployment of Common Data Elements for tissue banks for translational research in cancer—an emerging standard based approach for the Mesothelioma Virtual Tissue Bank. *BMC Cancer* 2008;8:91. Retrieved from http://w09.biomedcentral.com/content/pdf/1471-2407-8-91.pdf.

Morris MJ, Basch EM, Wilding G, et al. Department of the Defense Prostate Cancer Clinical Trials Consortium: a new instrument for prostate cancer clinical research. *Clin Genitourin Cancer.* 2009;7:51–57.

Morrison FP, Li L, Lai AM, et al. Repurposing the clinical record: can an existing natural language processing system de-identify clinical notes?. *J Am Med Inform Assoc.* 2009;16:37–39.

Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc.* 2010;17:124–130.

Nazi KM, Hogan TP, Wagner TH, et al. Embracing a health services research perspective on personal health records: lessons learned from the VA My HealtheVet system. *J Gen Intern Med.* 2010;25:62–67.

Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. *Methods Inf Med.* 2009;48(1):45–54.

Pace WD, Cifuentes M, Valuck RJ, et al. An electronic practice-based network for observational comparative effectiveness research. *Ann Intern Med.* 2009;151:338–340.

Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. *J Med Internet Res.* 2007;9:e15. Retrieved from http://www.jmir.org/2007/2/e15/.

Pakhomov S, Weston SA, Jacobsen SJ, et al. Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care.* 2007;13:281–288.

Pathak J, Peters L, Chute CG, et al. Comparing and evaluating terminology services application programming interfaces: RxNav, UMLSKS and LexBIG. *J Am Med Inform Assoc.* 2010;17:714–719.

Pathak J, Solbrig HR, Buntrock JD, et al. LexGrid: a framework for representing, storing, and querying biomedical terminologies from simple to sublime. *J Am Med Inform Assoc.* 2009;16:305–315.

Piwowar HA, Becich MJ, Bilofsky H, et al. Towards a data sharing culture: recommendations for leadership from academic health centers. *PLoS Med.* 2008;5:e183.

Platt R, Davis R, Finkelstein J, et al. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on Therapeutics. *Pharmacoepidemiol Drug Saf.* 2001;10:373–377.

Platt R, Wilson M, Chan KA, et al. The new Sentinel Network—improving the evidence of medical-product safety. *N Engl J Med.* 2009;361:645–647.

Quantin C, Bouzelat H, Allaert FA, et al. How to ensure data security of an epidemiological follow-up: quality assessment of an anonymous record linkage procedure. *Int J Med Inform.* 1998;49:117–122.

Rubin D, Napel S. Imaging informatics: toward capturing and processing semantic information in radiology images. *Yearb Med Inform.* 2010:34–42.

Safran C, Bloomrosen M, Hammond WE, et al. Expert Panel. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc.* 2007;14:1–9.

Sarkar IN. Biomedical informatics and translational medicine. *J Transl Med.* 2010;8:22. Retrieved from http://www.translational-medicine.com/content/pdf/1479-5876-8-22.pdf.

Saver B. One system for electronic health records. *Health Aff (Millwood).* 2010;29:1273.

Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc.* 2010;17:507–513.

Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf.* 2010;19:858–868.

Selker HP, Strom BL, Ford DE, et al. White paper on CTSA consortium role in facilitating comparative effectiveness research: September 23, 2009 CTSA consortium strategic goal committee on comparative effectiveness research. *Clin Transl Sci.* 2010;3:29–37.

Sharma A, Pan T, Cambazoglu BB, et al. VirtualPACS—a federating gateway to access remote image data resources over the grid. *J Digit Imaging.* 2009;22:1–10

Sibanda T, He T, Szolovits P, et al. Syntactically-informed semantic category recognition in discharge summaries. *AMIA Annu Symp Proc.* 2006:714–778.

Smedley D, Schofield P, Chen CK, et al. Finding and sharing: new approaches to registries of databases and services for the biomedical sciences. *Database.* 2010:baq014.

Solt I, Tikk D, Gál V, et al. Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier. *J Am Med Inform Assoc.* 2009;16:580–584.

Soto CM, Kleinman KP, Simon SR. Quality and correlates of medical record documentation in the ambulatory care setting. *BMC Health Serv Res.* 2002;2:22. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC140026/pdf/1472-6963-2-22.pdf.

Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med.* 2009;151:203–205.

Staes CJ, Xu W, LeFevre SD, et al. A case for using grid architecture for state public health informatics: the Utah perspective. *BMC Med Inform Dec Mak.* 2009;9:32. Retrieved from http://www.biomedcentral.com/content/pdf/1472-6947-9-32.pdf.

Stein HD, Nadkarni P, Erdos J, et al. Exploring the degree of concordance of coded and textual data in answering clinical queries from a clinical data repository. *J Am Med Inform Assoc.* 2000;7:42–54.

Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ.* 2009;338:b81. Retrieved from http://www.bmj.com/highwire/filestream/377160/field_highwire_article_pdf/0.pdf.

Tripathi M, Delano D, Lund B, et al. Engaging patients for health information exchange. *Health Aff (Millwood).* 2009;28:435-443.

(*Continued*)

Appendix. Summary Table of Articles on CER and Clinical Informatics Selected for Review and Discussion (*continued*)

Turisco F, Keogh D, Stubbs C, et al. Current status of integrating information technologies into the clinical research enterprise within US academic health centers: strategic value and opportunities for investment. *J Investig Med.* 2005;53:425–433.

Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. *J Am Med Inform Assoc.* 2007;14(5):550–563.

Uzuner O, Mailoa J, Ryan R, et al. Semantic relations for problem-oriented medical records. *Artif Intell Med.* 2010;50:63–73.

Uzuner O. Recognizing obesity and comorbidities in sparse data. *J Am Med Inform Assoc.* 2009;16:561–570.

Vardaki M, Papageorgiou H, Pentaris F. A statistical metadata model for clinical trials data management. *Comput Methods Prog Biomed.* 2009;95: 129–145.

Viangteeravat T, Brooks IM, Smith EJ, et al. Slim-prim: a biomedical informatics database to promote translational research. *Perspect Health Inf Manag.* 2009;6:6. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682606/pdf/phim0006-0006.pdf.

Ware H, Mullett CJ, Jagannathan V. Natural language processing framework to assess clinical conditions. *J Am Med Inform Assoc.* 2009;16:585–589.

Warzel DB, Andonaydis C, McCurry B, et al. Common data element (CDE) management and deployment in clinical trials. *AMIA Annu Symp Proc.* 2003;1048.

Weber GM, Murphy SN, McMurry AJ, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc.* 2009;16:624–630.

Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? *Ann Intern Med.* 2009;151:359–360.

Wellner B, Huyck M, Mardis S, et al. Rapidly retargetable approaches to de-identification in medical records. *J Am Med Inform Assoc.* 2007;14:564–573.

Weng C, Gennari JH, Fridsma DB. User-centered semantic harmonization: a case study. *J Biomed Inform.* 2007;40:353–364.

Wilcox A, Natarajan K, Weng C. Using personal health records for automated clinical trials recruitment: the ePaIRing model. *Summit on Translat Bioinforma.* 2009;2009:136–140.

Wilcox AB, Vawdrey DK, Chen YH, et al. The evolving use of a clinical data repository: facilitating data access within an electronic medical record. *AMIA Annu Symp Proc.* 2009;2009:701–705.

CER indicates comparative effectiveness research.

REFERENCES

1. US Department of Health and Human Services. Text of the Recovery Act Related to Comparative Effectiveness Funding. Excerpt from the American Recovery and Reinvestment Act of 2009. 2009. Available at: http://www.hhs.gov/recovery/programs/cer/recoveryacttext.html. Accessed August 31, 2011.

2. National Library of Medicine.New Coordinated Search Queries for PubMed®, Health Services Research Projects in Progress (HSRProj), and ClinicalTrials.gov. 2010. Available at: http://www.nlm.nih.gov/pubs/techbull/ma10/ma10_cer_nichsr.html. Accessed March 8, 2012.

3. PubMed search strategies applied October 2010.

4. National Library of Medicine.What is PubMed®? 2011. Available at: http://www.nlm.nih.gov/services/pubmed.html. Accessed March 8, 2012.

5. Murphy EK. *Secondary Use of HIT Data for Health Services Research Annotated Bibliography*. Washington, DC: AcademyHealth; 2010.

6. National Cancer Institute.caBIG Program. 2012. Available at: http://cabig.cancer.gov/. Accessed March 8, 2012.

7. AAFP National Research Network. DARTNet. 2012. Available at: http://www.dartnet.info. Accessed March 8, 2012.

8. Agency for Healthcare Research and Quality.DEcIDE Project Research Summaries, Reviews, and Reports. 2012. Available at: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?category=&search=&methodCategory=&agencyType=3&language=1&searchButton=Search. Accessed March 8, 2012.

9. HMO Research Network.HMO Research Network (HMORN). 2012. Available at: http://www.hmoresearchnetwork.org/. Accessed March 8, 2012.

10. University of California San Diego. iDASH. 2012. Available at: http://idash.ucsd.edu/. Accessed March 8, 2012.

11. Partners Healthcare.i2b2 Publications 2005–2011. 2012. Available at: https://www.i2b2.org/pubs/. Accessed March 8, 2012.

12. Foundation for the National Institutes of Health.OMOP Publications. 2011. Available at: http://omop.fnih.org/PublicationList. Accessed March 8, 2012.

13. Case Western Reserve University.Physio-MIMI Publications. 2012. Available at: http://physiomimi.case.edu/publicwiki/index.php/Publications. Accessed March 8, 2012.

14. Vanderbilt University.REDCap. 2012. Available at: http://project-redcap.org/. Accessed March 8, 2012.

15. Mini-Sentinel Coordinating Center.Mini-Sentinel Publications and Selected Presentations. 2011. Available at: http://mini-sentinel.org/publications/. Accessed March 8, 2012.

16. Office of the National Coordinator for Health Information Technology.Strategic Health IT Advanced Research Projects (SHARP) Program. 2011. Available at: http://healthit.hhs.gov/portal/server.pt/ community/healthit_hhs_gov__sharp_program/. Accessed February 22, 2012.

17. The Ohio State University.TRIAD Community. 2012. Available at: http://triadcommunity.org/. Accessed March 8, 2012.

18. US Department of Veterans Affairs.VA Informatics and Computing Infrastructure (VINCI). 2012. Available at: http://www.hsrd.research.va.gov/for_researchers/vinci/. Accessed February 22, 2012.

19. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med.* 2009;151: 203–205.

20. D'Avolio LW, Farwell WR, Fiore LD. Comparative effectiveness research and medical informatics. *Am J Med.* 2010;123(suppl 1):e32–e37.

21. Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. *J Am Med Inform Assoc.* 2009;16:316–327.

22. Corn M, Rudzinski KA, Cahn MA. Bridging the gap in medical informatics and health services research: workshop results and next steps. *J Am Med Inform Assoc.* 2002;9:140–143.

23. McCray AT, Scherrer JR, Safran C, et al. Concepts, knowledge, and language in health-care information systems. *Methods Inf Med.* 1995;34:1–4.

24. de Carvalho EC, Batilana AP, Simkins J, et al. Application description and policy model in collaborative environment for sharing of information on epidemiological and clinical research data sets. *PLoS One.* 2010;5:e9314.

25. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Fam Pract.* 2006;23:253–263.

26. Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. *Health Aff (Millwood).* 2009; 28:454–466.

27. Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. *Methods Inf Med.* 2009;48:45–54.

28. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. *J Med Internet Res.* 2007;9:e15. Retrieved from http://www.jmir.org/2007/2/e15/.

29. Piwowar HA, Becich MJ, Bilofsky H, et al. Towards a data sharing culture: recommendations for leadership from academic health centers. *PLoS Med.* 2008;5:e183.

30. Sarkar IN. Biomedical informatics and translational medicine. *J Transl Med.* 2010;8:22. Retrieved from http://www.translational-medicine.com/content/pdf/1479-5876-8-22.pdf.

31. Weber GM, Murphy SN, McMurry AJ, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query

tool for clinical data repositories. *J Am Med Inform Assoc.* 2009; 16:624–630.

32. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? *Ann Intern Med.* 2009;151:359–360.

33. McKinney M. 'Huge potential' for EHRs and comparative effectiveness. *Hosp Health Netw.* 2010;84:41–42.

34. Etheredge LM. Creating a high-performance system for comparative effectiveness research. *Health Aff (Millwood).* 2010;29:1761–1767.

35. Mandl KD, Lee TH. Integrating medical informatics and health services research: the need for dual training at the clinical health systems and policy levels. *J Am Med Inform Assoc.* 2002;9:127–132.

36. Buetow KH. An infrastructure for interconnecting research institutions. *Drug Discov Today.* 2009;14:605–610.

37. Fridsma DB, Evans J, Hastak S, et al. The BRIDG project: a technical report. *J Am Med Inform Assoc.* 2008;15:130–137.

38. Langella S, Hastings S, Oster S, et al. Sharing data and analytical resources securely in a biomedical research Grid environment. *J Am Med Inform Assoc.* 2008;15:363–373.

39. Manion FJ, Robbins RJ, Weems WA, et al. Security and privacy requirements for a multi-institutional cancer research data grid: an interview-based study. *BMC Med Inform Decis Mak.* 2009;9:31. Retrieved from http://www.biomedcentral.com/content/pdf/1472-6947-9-31.pdf.

40. Pace WD, Cifuentes M, Valuck RJ, et al. An electronic practice-based network for observational comparative effectiveness research. *Ann Intern Med.* 2009;151:338–340.

41. Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, et al. Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. *Med Care.* 2009;47:1091–1097.

42. Behrman RE, Benner JS, Brown JS, et al. Developing the Sentinel system—a national resource for evidence development. *N Engl J Med.* 2011;364:498–499.

43. Platt R, Davis R, Finkelstein J, et al. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on Therapeutics. *Pharmacoepidemiol Drug Saf.* 2001;10:373–377.

44. Chan KA. Development of a multipurpose dataset to evaluate potential medication errors in ambulatory settings and methodology. In: Henriksen KBJ, Marks ES, Lewin DI, eds. *Advances in Patient Safety: From Research to Implementation Vol 2: Concepts and Methodology. 2011/01/21 ed.* Rockville: Agency for Healthcare Research and Quality; 2005:225–238.

45. Deshmukh VG, Meystre SM, Mitchell JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. *BMC Med Res Methodol.* 2009;9:70. Retrieved from http://www.biomedcentral.com/content/pdf/1471-2288-9-70.pdf.

46. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc.* 2010;17:124–130.

47. Himes BE, Dai Y, Kohane IS, et al. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *J Am Med Inform Assoc.* 2009;16:371–379.

48. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42:377–381.

49. Platt R, Wilson M, Chan KA, et al. The new Sentinel Network—improving the evidence of medical-product safety. *N Engl J Med.* 2009;361:645–647.

50. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf.* 2010;19:858–868.

51. Hastings S, Oster S, Langella S, et al. Adoption and adaptation of caGrid for CTSA. *Summit Translat Bioinform.* 2009:44–48.

52. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform.* 2009;42:760–772.

53. Ambert KH, Cohen AM. A system for classifying disease comorbidity status from medical discharge summaries using automated hotspot and negated concept detection. *J Am Med Inform Assoc.* 2009;16:590–595.

54. Bramsen P, Deshpande P, Lee YK, et al. Finding temporal order in discharge summaries. *AMIA Annu Symp Proc.* 2006;81–85.

55. Carrell D, Miglioretti D, Smith-Bindman R. Coding free text radiology reports using the Cancer Text Information Extraction System (caTIES). *AMIA Annu Symp Proc.* 2007;889.

56. Clark C, Good K, Jezierny L, et al. Identifying smokers with a medical extraction system. *J Am Med Inform Assoc.* 2008;15:36–39.

57. Elkins JS, Friedman C, Boden-Albala B, et al. Coding neuroradiology reports for the Northern Manhattan Stroke Study: a comparison of natural language processing and manual review. *Comput Biomed Res.* 2000;33:1–10.

58. Farkas R, Szarvas G, Hegedus I, et al. Semi-automated construction of decision rules to predict morbidities from clinical texts. *J Am Med Inform Assoc.* 2009;16:601–605.

59. Friedlin J, Grannis S, Overhage JM. Using natural language processing to improve accuracy of automated notifiable disease reporting. *AMIA Annu Symp Proc.* 2008;207–211.

60. Friedman C, Shagina L, Lussier Y, et al. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc.* 2004;11:392–402.

61. Goryachev S, Kim H, Zeng-Treitler Q. Identification and extraction of family history information from clinical reports. *AMIA Annu Symp Proc.* 2008;247–251.

62. Hazlehurst B, Naleway A, Mullooly J. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. *Vaccine.* 2009;27:2077–2083.

63. Hazlehurst B, Sittig DF, Stevens VJ, et al. Natural language processing in the electronic medical record: assessing clinician adherence to tobacco treatment guidelines. *Am J Prev Med.* 2005;29: 434–439.

64. Honigman B, Lee J, Rothschild J, et al. Using computerized data to identify adverse drug events in outpatients. *J Am Med Inform Assoc.* 2001;8:254–266.

65. Liao KP, Cai T, Gainer V, et al. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken).* 2010;62:1120–1127.

66. McCormick PJ, Elhadad N, Stetson PD. Use of semantic features to classify patient smoking status. *AMIA Annu Symp Proc.* 2008;450–454.

67. Meystre SM, Friedlin FJ, South BR, et al. Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC Med Res Methodol.* 2010;10:70. Retrieved from http://www.biomedcentral.com/content/pdf/1472-6947-8-60.pdf.

68. Mishra NK, Cummo DM, Arnzen JJ, et al. A rule-based approach for identifying obesity and its comorbidities in medical discharge summaries. *J Am Med Inform Assoc.* 2009;16:576–579.

69. Morrison FP, Li L, Lai AM, et al. Repurposing the clinical record: can an existing natural language processing system de-identify clinical notes? *J Am Med Inform Assoc.* 2009;16:37–39.

70. Pakhomov S, Weston SA, Jacobsen SJ, et al. Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care.* 2007;13(pt 1):281–288.

71. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc.* 2010;17: 507–513.

72. Sibanda T, He T, Szolovits P, et al. Syntactically-informed semantic category recognition in discharge summaries. *AMIA Annu Symp Proc.* 2006;714–718.

73. Solt I, Tikk D, Gal V, et al. Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier. *J Am Med Inform Assoc.* 2009;16:580–584.

74. Uzuner O, Mailoa J, Ryan R, et al. Semantic relations for problem-oriented medical records. *Artif Intell Med.* 2010;50:63–73.

75. Chute CG, Elkin PL, Sherertz DD, et al. Desiderata for a clinical terminology server. *Proc AMIA Symp.* 1999;42–46.

76. Ware H, Mullett CJ, Jagannathan V. Natural language processing framework to assess clinical conditions. *J Am Med Inform Assoc.* 2009;16:585–589.

77. Wellner B, Huyck M, Mardis S, et al. Rapidly retargetable approaches to de-identification in medical records. *J Am Med Inform Assoc.* 2007;14:564–573.

78. Wilcox AB, Vawdrey DK, Chen YH, et al. The evolving use of a clinical data repository: facilitating data access within an electronic medical record. *AMIA Annu Symp Proc*. 2009;2009:701–705.

79. Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. *J Am Med Inform Assoc*. 2007;14: 550–563.

80. Uzuner O. Recognizing obesity and comorbidities in sparse data. *J Am Med Inform Assoc*. 2009;16:561–570.

81. Neches R, Fikes RE, Finin T, et al. Enabling technology for knowledge sharing. *AI Magazine*. 1991;12:16–36.

82. Chute CG, Beck SA, Fisk TB, et al. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *J Am Med Inform Assoc*. 2010;17:131–135.

83. Chute CG. Clinical classification and terminology: some history and current observations. *J Am Med Inform Assoc*. 2000;7:298–303.

84. Deitzer JR, Payne PR, Starren JB. Coverage of clinical trials tasks in existing ontologies. *AMIA Annu Symp Proc*. 2006:903.

85. Pathak J, Solbrig HR, Buntrock JD, et al. LexGrid: a framework for representing, storing, and querying biomedical terminologies from simple to sublime. *J Am Med Inform Assoc*. 2009;16:305–315.

86. Rubin DL, Napel S. Imaging informatics: toward capturing and processing semantic information in radiology images. *Yearb Med Inform*. 2010;34–42.

87. Weng C, Gennari JH, Fridsma DB. User-centered semantic harmonization: a case study. *J Biomed Inform*. 2007;40:353–364.

88. Morris MJ, Basch EM, Wilding G, et al. Department of defense prostate cancer clinical trials consortium: a new instrument for prostate cancer clinical research. *Clin Genitourin Cancer*. 2009;7:51–57.

89. Mohanty SK, Mistry AT, Amin W, et al. The development and deployment of Common Data Elements for tissue banks for translational research in cancer—an emerging standard based approach for the Mesothelioma Virtual Tissue Bank. *BMC Cancer*. 2008;8:91. Retrieved from http://w09.biomedcentral.com/content/pdf/1471-2407-8-91.pdf.

90. Warzel DB, Andonaydis C, McCurry B, et al. Common data element (CDE) management and deployment in clinical trials. *AMIA Annu Symp Proc*. 2003;1048.

91. Kukafka R, Ancker JS, Chan C, et al. Redesigning electronic health record systems to support public health. *J Biomed Inform*. 2007;40: 398–409.

92. Hynes DM, Perrin RA, Rappaport S, et al. Informatics resources to support health care quality improvement in the veterans health administration. *J Am Med Inform Assoc*. 2004;11:344–350.

93. Saver B. One system for electronic health records. *Health Aff (Millwood)*. 2010;29:1273.

94. Kush RD, Helton E, Rockhold FW, et al. Electronic health records, medical research, and the Tower of Babel. *N Engl J Med*. 2008;358: 1738–1740.

95. Brown JS, Holmes JH, Shah K, et al. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care*. 2010;48(suppl):S45–S51.

96. Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. *BMC Public Health*. 2006;6:235. Retrieved from http://www.biomedcentral.com/content/pdf/1471-2458-6-235.pdf.

97. Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. *Ann Intern Med*. 2009;151:341–344.

98. Carpenter PC, Chute CG. The Universal Patient Identifier: a discussion and proposal. *Proc Annu Symp Comput Appl Med Care*. 1993:49–53.

99. El Emam K, Jabbouri S, Sams S, et al. Evaluating common de-identification heuristics for personal health information. *J Med Internet Res*. 2006;8:e28. Retrieved from http://www.jmir.org/2006/4/e28/.

100. Fefferman NH, O'Neil EA, Naumova EN. Confidentiality and confidence: is data aggregation a means to achieve both? *J Public Health Policy*. 2005;26:430–449.

101. Krishna R, Kelleher K, Stahlberg E. Patient confidentiality in the research use of clinical medical databases. *Am J Public Health*. 2007;97:654–658.

102. McGraw D, Dempsey JX, Harris L, et al. Privacy as an enabler, not an impediment: building trust into health information exchange. *Health Aff (Millwood)*. 2009;28:416–427.

103. Black N. Secondary use of personal data for health and health services research: why identifiable data are essential. *J Health Serv Res Policy*. 2003;8(suppl 1):36–40.

104. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and ethical considerations for linking clinical and administrative databases. *Am Heart J*. 2009;157:971–982.

105. Goldstein MM. Health information technology and the idea of informed consent. *J Law Med Ethics*. 2010;38:27–35.

106. Sabharwal R, Holve E, Rein A, et al. *Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy*. Washington, DC: AcademyHealth EDM Forum Issue Brief). Available at: http://www.edm-forum.org; 2012.

107. Berger M. *Health Care, Cost, Quality, and Outcomes: ISPOR Book of Terms*. Lawrenceville, NJ: International Society for Pharmacoeconomics and Outcomes Research; 2003.