Published in final edited form as:

Curr Oncol Rep. 2012 December ; 14(6): 494–501. doi:10.1007/s11912-012-0272-6.

Leveraging EHR Data for Outcomes and Comparative Effectiveness Research in Oncology

Frank J. Manion, MS,

University of Michigan Comprehensive Cancer Center, 1600 Huron Parkway, SPC 2800, Ann Arbor, MI 48109-2800, 734-764-2473 Phone, 734-998-6155 Fax

Marcelline R. Harris, PhD, RN,

University of Michigan School of Nursing, 400 North Ingalls, RM 4160, Ann Arbor MI 48109-5482, 734/763-4995 Phone

Ayşe G. Buyuktur, MPH, MS,

University of Michigan School of Information, 105 South State Street, Ann Arbor MI 48104, 734-763-2285 Phone

Patricia M. Clark, PhD, RN,

University of Michigan Comprehensive Cancer Center, 300 North Ingalls, RM 8C29, Ann Arbor MI 48109-5473, 734-647-8349 Phone

Lawrence C. An, MD, and

University of Michigan Comprehensive Cancer Center, 300 N Ingalls, RM 5D04, Ann Arbor MI 48109-0471, 734-763-6099 Phone

David A. Hanauer, MD, MS

University of Michigan Comprehensive Cancer Center, 1600 Huron Parkway, SPC 5456, Ann Arbor, MI 48109-2800, 734/615-0599 Phone

Frank J. Manion: fmanion@umich.edu; Marcelline R. Harris: mrhrrs@umich.edu; Ayşe G. Buyuktur: abuyuktu@umich.edu; Patricia M. Clark: clarkpat@umich.edu; Lawrence C. An: lcan@umich.edu; David A. Hanauer: hanauer@umich.edu

Abstract

Along with the increasing adoption of electronic health records (EHRs) are expectations that data collected within EHRs will be readily available for outcomes and comparative effectiveness research. Yet the ability to effectively share and reuse data depends on implementing and configuring EHRs with these goals in mind from the beginning. Data sharing and integration must be planned both locally as well as nationally. The rich data transmission and semantic infrastructure developed by the National Cancer Institute (NCI) for research provides an excellent example of moving beyond paper-based paradigms and exploiting the power of semantically robust, network-based systems, and engaging both domain and informatics expertise. Similar efforts are required to address current challenges in sharing EHR data.

Keywords

Data sharing; Comparative effectiveness research; Outcomes research; Clinical informatics; Common data elements; Semantic interoperability; Electronic health records

Correspondence to: Frank J. Manion, fmanion@umich.edu.

Disclosure

No potential conflicts of interest relevant to this article were reported.

Introduction

Data sharing is widely recognized as essential to advancing health and the delivery of health care services. The data in electronic health records (EHRs) are of particular interest, because these data can enable new insights into naturally occurring variations in disease processes, response to treatments, and patient care delivery. Expectations are that investments in health information technology, particularly EHRs, will expand the potential to collect and subsequently share data. Recent Institute of Medicine (IOM) reports highlight the tremendous opportunity to improve health outcomes when the explosion of data—from the molecular level to patient and population levels—can be shared [1, 2]. Within the oncology community, the National Cancer Institute (NCI) has long focused on infrastructure components required to enable data sharing across networked environments [3].

The federal government's recognition of the need to stimulate the adoption of EHRs is evident in the American Recovery and Reinvestment Act (ARRA), specifically the Health Information Technology for Economic and Clinical Health (HITECH) Act. HITECH incentivizes the implementation of EHRs and their meaningful use. Even with the increasing adoption of certified EHRs, a gap remains in the ability to share EHR data. Data sharing is dependent on a well-coordinated set of technologies, processes, and governance. In this paper, we focus on a set of technical issues that, if addressed up-front at the time EHRs are implemented, makes the downstream sharing of data much easier. Definitions of key terms and abbreviations are provided at the end of the article.

EHRs as a Source of Data for Outcomes and Comparative Effectiveness Research

Electronic health records are primarily designed as point-of-care information systems that support the delivery of patient care in the context of a patient encounter with a provider. EHRs were not designed for secondary uses, such as outcomes and comparative effectiveness research. Rather, EHR data must be retrieved from the transactional EHR database and then linked to patient-level data from other sources, including clinical trials, cohorts and registries, imaging, etc. [4]. A review of the literature through 2006 reported that more than half of the 126 studies using EHRs for outcomes research supplemented the EHR data with other data, either patient reported (40 %), paper-based chart data (30 %), or pharmacy/lab data (17 %) [5].

Because transactional systems are not optimized for data retrieval, data are typically exported from the transactional database and linked to data from other source systems in repositories specifically purposed for analysis (e.g., data warehouses). The challenges of using data from heterogeneous sources are increasingly recognized and include selection bias, varying timeframes for updating and linking data from different source systems, lack of agreement on common data elements or coding systems, and local variations in naming and coding data elements. [6, 7, 8]. Examples of promising approaches to overcome these challenges are evident in the electronic medical records and genomics (eMERGE) network; a collaboration that is investigating the use of EHR data linked to DNA biorepositories to identify specific associations between genotypes and human phenotypes [9, 10]. Unfortunately, leveraging data for secondary use purposes is infrequently considered in the development, purchase or installation of source systems, including EHRs.

Opportunities to share EHR data across large, multi-institutional initiatives have created requirements to address a wide range of topics in the context of distributed or networked environments. Issues that were well understood in the era of paper-based clinical records that must now be re-examined in the context of sharing EHR data include: (a) the protection

of the security and privacy of individual data; (b) the notion of episodes and longitudinal perspectives as organizing constructs for an EHR; (c) accuracy and completeness of EHR data; (d) regulatory frameworks; and (e) assurances that principles of responsible conduct of research will be upheld. In addition, the need for technologies that enable software to “understand” and act on the data being transmitted in a digital format, known as semantic interoperability, has emerged as one of the most fundamental issues for effectively sharing health care data. While an in-depth discussion of all the technical issues involved in data sharing and semantic interoperability is beyond the scope of this paper, an extensive discussion in the context of cancer research can be found in a recent book by Ochs and colleagues [11•].

Semantic Interoperability

Semantic interoperability refers to “the ability of two or more information systems to communicate information and have that information properly interpreted by a receiving system in the same sense as was intended by the transmitting system” [12]. For example, the data element “5” carries with it no information that enables software or query tools to understand whether that data element represents an age (let alone if that age is in years, months, days or hours), a floor in a hospital or clinic, or any number of other concepts for which “5” may be a legitimate value. Standardized approaches to representing the semantics of data elements are needed for dynamic, run-time matching of data coming from different source systems, for automated translation of content when sending messages, and for concept-based query tools.

Sharing a concern about the need for integrated and scalable solutions for semantic interoperability in health care, several groups are beginning to consider “interoperability frameworks.” For example, Health Level 7 (HL7), a major standards development organization, has expanded its mission beyond messaging standards, and is addressing “net-centric” data standards and distributed computing requirements for interoperability [12]. In early 2012, the Office of the National Coordinator (ONC) launched an effort to engage volunteers from the public and private sector to work on a Standards and Interoperability (S&I) Framework [13]. Semantic interoperability is important for dealing with heterogeneity across data sources; in EHRs the issue is not yet sufficiently addressed.

The Role of Standards in Semantics and Data Sharing

Standards operate across multiple levels. Some standards, while extremely important for sharing data in a networked environment, exist at purely technical levels of data interchange. Examples include XML, HTML, HTTP, and TCP/IP, which are involved in structuring, displaying and transporting data between systems. These standards are largely independent of the underlying meaning of the data being shared. As an example, when a web server provides a web page to a browser, the browser will render and display whatever is on the web page, regardless of its content or meaning.

Other standards focus on common data elements (CDEs) (see definitions) and reflect a community-based consensus about terms, definitions, and allowable values [14]. Notable examples of CDEs in the area of cancer research include those developed by the American Joint Committee on Cancer (AJCC), Surveillance, Epidemiology and End Results (SEER), and North American Association of Central Cancer Registries (NAACCR). These groups have all developed standards for data coding and data transmission to state and national-level cancer registries, although earlier methods for exchanging these standard terms, definition and values was typically done in codebooks or flat files such as spreadsheets. The NCI, to address newer web-based technologies, developed a registry based on international technical standards for storing not only the CDEs, but also “metadata” (i.e., descriptive

information about each data element), known as the Cancer Data Standards Repository (caDSR) [15]. This metadata standard (referred to by its technical name of ISO 11179) is based on semantic theory and requires that: (a) data elements are named and defined; (b) data elements representing the same concept (e.g., country code and country name) are assigned a common concept code; (c) allowable values for each concept are enumerated to avoid ambiguity; and (d) conceptual domains are described. In addition to data interchange and CDE-focused standards, achieving the level of semantic interoperability required for EHR data sharing requires that standard terminology (e.g., Systematized Nomenclature of Medicine--Clinical Terms (SNOMED-CT) and Logical Observation Identifiers Names and Codes (LOINC)) and abstract models of information can be exchanged (e.g., caBIG and the Clinical Data Interchange Standards Consortium (CDISC)) [16].

Within the oncology community, there are several examples where standards-based approaches to semantic interoperability have been successfully used for multi-site distributed or virtual research groups. 'A Growable Network Information System' (AGNIS) of the Center for International Blood and Marrow Transplant Research (CIBMTR) is built on such a framework, and includes a use of the caBIG information model and caDSR CDEs developed by the CIBMTR [17–20]. AGNIS is used to exchange hematopoietic cell transplant data in a standardized, unambiguous manner so that data from multiple institutions can be seamlessly shared, merged, and re-used.

While methods enabling the semantic interoperability of data are developed and implemented in some research settings, their widespread adoption in EHRs is limited. This may be partly due to the lack of financial incentives for EHR vendors to adopt standards. The recent Meaningful Use financial incentives may influence standards adoption, but it remains to be seen if EHR vendors are willing or even able to turn around their legacy technologies and fully embrace standards for semantic interoperability. As clinicians become familiar with the issues around technical aspects of data sharing, there are opportunities to influence “up front” decisions surrounding EHR implementations that could improve the likelihood of data sharing.

Practical Considerations for EHRs in the Context of Research

When data are entered into an EHR, those data should be available in the underlying database for extraction, sharing, and secondary use. However, situations have arisen in which patient records were available electronically, but useful information was not easily obtainable. This was described in a recent paper with the apt title, “Instant availability of patient records, but diminished availability of patient information” [21]. The ease with which data can be extracted from an EHR for secondary use and sharing is heavily influenced by how data were initially entered. The two most basic forms of data entry are structured and unstructured. Structured data entry is often achieved via drop down menus, check boxes, or other ‘point and click’ features. Free text is often created via typing or dictation/transcription. Various EHRs may support both structured and unstructured data entry and hybrid models.

In general, high quality coded data are much easier to use than free text data, although not all data are equally valuable or reliable. For example, ICD-9 billing codes are often easily accessible, but their value for research is challenged [22–24]. Free text is often richer in detail but requires specialized approaches, including natural language processing (NLP) [25] to extract clinically relevant concepts, such as those in cancer treatment summaries [26]. Other options include search engines supporting human-driven data abstraction [27], although these are not widely available. Regardless of which approach is taken—structured or free-text with NLP methods applied to “elementize” unstructured data—if data are to be

shared and aggregated, then the importance of the ultimate terminology coding and use of information models cannot be overstated. It is ultimately these frameworks that allow effective integration between systems at the semantic level.

Differences of opinion exist between clinicians who value one approach for data entry over another. Some clinicians favor structured data entry because timesaving templates can be used that pre-populate data from other sections of the EHR, including laboratory data or medication lists. However, structured templates often impose constraints on the way data are entered, and this may not support 'natural' thought processes and workflows for data entry and communication. Free text can sometimes be faster to enter, especially for complex oncology cases, and the more descriptive narratives may aid in communication with other clinicians or patients.

The tension that arises when making decisions between free text versus structured documentation remains an open issue [28]. In situations where it is known that data will need to be re-used or shared, efforts should be made to enter the data in a coded manner; however, it is important to understand the potential trade-offs. While a recent study reported that clinicians who used structured EHR documentation had better 'quality of care' than their colleagues who dictated their notes [29], other studies have come to different conclusions [30]. Some clinicians have found that structured, coded data entry approaches often result in documents that are very generic, lacking details that are important in distinguishing one patient from another [31]. Hybrid solutions that combine structured data entry with options for free text components are possible, but even in these situations, it is possible that structured data entry can be circumvented with users typing information that otherwise could have been coded [32].

Case Example

The issues and examples described above provide a basis for thinking about best practices when implementing EHRs, and the important implications about the choices made for documentation. Epic Systems Corporation (Verona, WI) is an EHR vendor that is quickly becoming one of the largest providers of software to health systems across the country. Although their market is primarily large medical centers, with less penetration into smaller office based practices, the company is growing at a rapid pace [33]. Thus, the functionality provided by Epic provides a useful window into understanding the possibilities, and challenges, of developing a system to support the secondary use and sharing of clinical data.

Various vendors take different approaches to implementing EHR systems. While some vendors provide detailed, pre-configured components and documentation templates, Epic does not. Rather, they provide a 'model system,' which essentially contain examples of what could be done in terms of a clinical build, but which most clinicians would not consider to be usable in a clinical environment. It is then up to each institution to develop their own clinical content, either through internal working groups, through direct interactions with other institutions, or via Epic User Groups. Local implementation decisions include choices not only about documentation structure, but about the use of specific terminology systems for features ranging from medication formulary lists (e.g., RxNorm), laboratory data (e.g. LOINC), to problem list elements (e.g., SNOMED-CT).

An advantage of this locally driven approach is that it allows each institution to customize the system to their specific needs. The disadvantage is that most institutions and practices don't necessarily know what their needs are until well after the implementation has occurred, and it can be very difficult among a large number of users to standardize clinical content across, or even within, disciplines and specialties. Usage patterns of EHRs can actually vary even within the same institution [34].

Based on our experience, it is important that all clinical groups put upfront effort into standardizing their processes for data collection, especially for data that are known to be important for secondary use and sharing. This may involve standardizing data elements on specific CDEs, terminologies, and defining an essential set of data elements which must be coded, and which can be captured as free text. Choosing a common location in the medical record is important, since some data elements might reasonably be recorded in multiple locations, making them harder to locate later. An example we recently encountered involved the recording of cancer staging information. After various meetings and input from clinical groups, it was decided to record staging in the problem list, so that it could easily be found and extracted by other clinicians. The alternative would have been to have staging ‘buried’ in the clinical notes, making it much harder to find.

Our experience also provides an example in the complexities of extracting data for secondary purposes. The underlying database in Epic, called Chronicles, is based on a system called Cache. Cache, in turn, is based on MUMPS (the Massachusetts General Hospital Utility Multi-Programming System), initially developed nearly a half-century ago. MUMPS is efficient but complex, and therefore, a subset of the Epic data are transferred nightly to a second database called Clarity. Clarity uses databases that support the structured query language (SQL), which is the current standard for extracting data. This is advantageous except for a few not-so-minor details. First, Clarity only represents a subset of what the full Chronicles database contains. As a result, some data may not be readily available. Second, because the data are not updated in real time, some uses of the data might not be possible—such as building a system that utilizes an up-to-the-minute patient schedule for identifying eligible study patients that have just arrived in the clinic. Third, Clarity contains thousands of database tables, making the extraction of data complex, even for talented SQL programmers. As a result, some institutions have had to build yet a third instance of Epic data using simplified data structures that are easier to understand and query.

Because of the complexity of the data and underlying system architectures, strong technical skills are required to extract the data. In fact, Epic often requires that individuals who extract data from Clarity become ‘certified’ or ‘proficient,’ by studying educational materials and subsequently passing an examination. As a result, extracting data should probably be done by informatics or information technology professionals.

Other challenges also exist in using the data that have been stored in Clarity. Much of the ‘structure’ of the documentation is stripped out, including line feeds and table grids. Thus metadata is lost, and documents that initially had a structured table of data or had information presented where line feeds matter, lose some integrity with this transformation. Bypassing this limitation often requires complex technical interventions, such as intercepting HL7 feeds as data are passed between systems and databases. Furthermore, even if data are structured it is still necessary to know the meaning of it; that is, what the stored codes actually represent. For example, if a clinical group chooses to classify pain on a scale of 1 to 5, then it will be necessary to know whether the 1 or the 5 represents the most severe pain. Such metadata might not be captured in the database, and therefore accurately recording those details in a separate searchable database may become important. The issue might be complicated further if another group choose the same scale for pain but with reversed meaning of the numbered sequence, or if a group chooses to measure pain on a completely different scale (e.g., “none”, “moderate”, “severe”). Complex mapping of concepts may be required to integrate data from disparate sources, even from different clinical groups within a single Epic installation. Each organization needs to manage terminology concepts in order to share data within an institution. Data sharing between institutions also requires management of terminology concepts, highlighting the importance of using standardized and concept-based terminologies (e.g., SNOMED-CT).

When clinical groups can agree on essential data elements, standardized definitions, and agree to adopt semantic interoperability standards, then sharing data and subsequent data extractions becomes easier. This can greatly reduce the amount of time needed long-term for continued data extraction and use. Data from one source (e.g. EHRs) could be automatically shared with other receiving systems that similarly agree to adopt semantic interoperability standards. The primary lesson to be learned is that dealing with the issues up-front, including the use of standards for semantic interoperability and data transfer, can make the downstream sharing of data much easier.

Future Directions

A question posed at a recent IOM workshop on informatics needs and challenges in cancer research exemplifies the current challenge facing EHRs: “Are we building systems and infrastructure that merely support the collection of data, or an integrated knowledge ecosystem that supports data validation, sophisticated analytics, evidence generation, and actionable knowledge to drive a learning healthcare enterprise?”[35].

As a community, oncology has been a leader in developing CDEs, terminology systems, knowledge models, and other technical infrastructure to support the level of semantic interoperability required for secondary uses such as research. The termination of the caBIG program [36] and its subsequent replacement with the National Cancer Informatics Program presents opportunity for ongoing development of interoperable biomedical information systems, built on community-driven data standards [37].

A 2010 report from the President’s Council of Advisors on Science and Technology [38] emphasized the need for such interoperability, and noted that current standards for exchanging data, vocabularies, and messages are not sufficient “to advance the state of the art either of clinical practice or of a robust health IT infrastructure.” The report called on the government and the community to develop a “universal exchange language,” focused on the technical ability to exchange data in uniform ways. New frameworks and models, similar to those used in other industries to inform large-scale integration and data sharing, are needed to guide scalable, multi-institutional use of EHR data [39].

Conclusion

Data sharing, and particularly the sharing of EHR data with other electronic systems, is becoming more feasible, as new technical frameworks and tools are being developed to assist with the challenges involved. However, fully realizing the potential of these newer approaches requires an upfront effort to standardize technical data interchange, data elements, terminology systems, metadata, and the information or domain models that describe the processes and context in which data are shared. Collaborations among clinicians, who understand the meaning of data within the domain, and informaticists, who understand the complexities of data sharing in networked environments, are essential to achieving success in this effort.

Acknowledgments

This project was supported (in part) by the National Institutes of Health through the University of Michigan’s Cancer Center Support Grant (5 P30 CA46592), and by the National Center for Research Resources (Award Number UL1RR024986). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Center for Research Resources.

Appendix: Definitions

caDSR is a cancer-specific set of common data elements and the metadata for cancer research developed through the caBIG project. Examples include preferred terms, definitions, mapping to reference coding systems etc. [15, 19, 20].

Common Data Elements (CDEs) are data elements and annotations defined as standards across research or clinical projects. CDEs and relationships with other data elements are typically maintained in a metadata repository. Effective CDEs are typically developed by multi-disciplinary groups and validated in successive rounds of critical analysis [14].

Comparative effectiveness research (CER) is the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care. CER aims to answer questions such as “What works best, for whom, and under what conditions?” [40, 41].

Data sharing refers to the set of technologies, standards, regulations, and trust factors that make data collected for one purpose electronically available for other purposes. Data reuse and secondary use of data are often used synonymously with data sharing.

Electronic health records (EHRs) are longitudinal electronic records of patients’ health information, generated by one or more encounters in any care delivery setting, and commonly include patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports [42].

Meaningful use (MU) refers to the Centers for Medicare and Medicaid Services (CMS) program that provides a financial incentive for the “meaningful use” of certified EHR technology. The American Recovery and Reinvestment Act of 2009 specifies three main components: (a) use of a certified EHR in a meaningful manner; (b) use of certified EHR technology for electronic exchange of health information; and (c) use of certified EHR technology to submit clinical quality and other measures [43].

Outcomes research encompasses a set of methodologies long used by the health services research community to study aspects of health care delivery [44]. Outcomes and effectiveness research are facilitated by the integration of large-scale, multi-institutional data from EHRs, clinical trial management systems, pharmacy, radiology, and disease registries such as SEER.

Structured data refers to data that are organized into a structure such as fixed fields, and often stored in a structured database organized by columns and rows. Structured data are often coded according to some agreed upon coding system such as ICD-9, SNOMED-CT, or AJCC TNM standards for coding tumor stage. Coded structured data can be made available for data sharing through a variety of interfaces including Web browsers, database query languages, application-specific interfaces, or data exchange formats.

Unstructured data refers to data such as free text that is not captured and stored in fixed fields, and not readily stored in rows and columns in databases. EHR documents such as physician notes and discharge summaries are largely composed of unstructured data. Increasingly, computer assisted techniques such as natural language processing (NLP) are being used to convert unstructured data to structured data—greatly reducing the need for manual chart review and subsequent data entry to convert unstructured to structured data.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1. Committee on a Framework for Development a New Taxonomy of Disease, National Research Council. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. The National Academies Press; 2011. Available at http://www.nap.edu/openbook.php?record_id=13284
2. Grossman, C.; Powers, B.; McGinnis, JM. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. Institute of Medicine; 2011.
3. Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu W-L, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform.* 2007; 40(1):30–43. [PubMed: 16697710]
4. Ohmann C, Kuchinke W. Future Developments of Medical Informatics from the Viewpoint of Networked Clinical Research. *Meth Inform Med.* 2009; 48(1):45–54. [PubMed: 19151883]
5. Dean BB, Lam J, Natoli JL, Butler Q, Aguilar D, Nordyke RJ. Review: Use of Electronic Medical Records for Health Outcomes Research. *Med Care Res Rev.* 2009; 66(6):611–638. [PubMed: 19279318]
6. Abernethy AP, Etheredge LM, Ganz PA, Wallace P, German RR, Neti C, et al. Rapid-Learning System for Cancer Care. *J Clin Oncol.* 2010; 28(27):4268–4274. [PubMed: 20585094]
7. Holve, E.; Segal, C.; Lopez, MH.; Rein, A.; Johnson, BH. [Accessed July 2012.] The Electronic Data Methods (EDM) Forum for Comparative Effectiveness Research (CER). Medical Care [Internet]. 2012. p. 50 Available from: http://journals.lww.com/lww-medicalcare/Fulltext/2012/07001/The_Electronic_Data_Methods_EDM_Forum_for.5.aspx
8. Kanas G, Morimoto L, Mowat F, O'Malley C, Fryzek J, Nordyke R. Use of electronic medical records in oncology outcomes research. *ClinicoEconomics and outcomes research: CEOR.* 2010; 2:1–14. This article discusses in depth the challenges to using EHR data in oncology, and includes a useful table on the advantages and disadvantages of a number of datasets used for outcomes research. [PubMed: 21935310]
9. Pathak J, Wang J, Kashyap S, Basford M, Li R, Masys DR, et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. *J Am Med Informat Assoc.* 2011; 18(4):376–386. This article describes in depth the process of using Common Data Elements to enable the integration and analysis of phenotypic data collected at multiple institutions as part of the Electronic Medical Records and Genomics (eMERGE) project; a project with the aim of correlating whole genome scans with phenotype data extracted from electronic health records.
10. Kho AN, Pacheco JA, Peissig PL, Rasmussen L, Newton KM, Weston N, et al. Electronic Medical Records for Genetic Research: Results of the eMERGE Consortium. *Science Translational Medicine.* 2011; 3(79):79re1.
11. Ochs, MF.; Casagrande, JT.; Davuluri, RV. Biomedical informatics for cancer research. New York: Springer; 2010. The first nine chapters of this recently published book provide in-depth coverage of the technologies and techniques discussed in this paper. The final eleven chapters discuss software tools relevant to the cancer research community
12. Landgrebe, J.; Smith, B. The HL7 Approach to Semantic Interoperability. Presented at the International Conference on Biomedical Ontology; Buffalo, New York. Available at <http://ontology.buffalo.edu/smith/articles/HL7-SAIF.pdf>
13. [Accessed July 2012.] Standards & Interoperability (S&I) Framework. Available at <http://wiki.siframework.org/>
14. Patel A, Kajdacsy-Balla A, Berman J, Bosland M, Datta M, Dhir R, et al. The development of common data elements for a multi-institute prostate cancer tissue bank: The Cooperative Prostate Cancer Tissue Resource (CPCTR) experience. *BMC Cancer.* 2005; 5(1):108. [PubMed: 16111498]

15. NCI Center for Biomedical Informatics and Information Technology. [Accessed July 2012.] Cancer Data Standards Registry and Repository (caDSR). Available at <https://cabig.nci.nih.gov/community/concepts/caDSR/>
16. CDISC. [Accessed July 2012.] Available at <http://www.cdisc.org/>
17. [Accessed July 2012.] A Growable Network Information System (AGNIS). Available at <http://www.agnis.net/>
18. [Accessed July 2012.] Center for International Blood & Marrow Transplant Research (CIBMTR) [Internet]. Available at <http://www.cibmtr.org/pages/index.aspx>
19. [Accessed June 2012.] caBIG. Available at <https://cabig.nci.nih.gov/>
20. Beck, JR. caBIG Strategic Planning Workspace. The Cancer Biomedical Informatics Grid(TM): Infrastructure and applications for a worldwide research community. Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems; 2007; 2007. p. 330-334.
21. Christensen T, Grimsø A. Instant availability of patient records, but diminished availability of patient information: A multi-method study of GP's use of electronic patient records. BMC Med Informat Decis Making. 2008; 8(1):12.
22. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. Med Care. 1998; 36(9):1324–36. [PubMed: 9749656]
23. Rhodes ET, Laffel LMB, Gonzalez TV, Ludwig DS. Accuracy of Administrative Coding for Type 2 Diabetes in Children, Adolescents, and Young Adults. Diabetes Care. 2007; 30(9):e99.
24. Leibson CL, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, et al. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008; 46(2):127–32. [PubMed: 18219240]
25. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Informat Assoc. 2011; 18(5):544–551.
26. Warner JL, Anick P, Hong P, Xue N. Natural language processing and the oncologic history: is there a match? J Oncol Pract. 2011; 7(4):e15–19. [PubMed: 22043196]
27. Zheng K, Mei Q, Hanauer DA. Collaborative search in electronic health records. J Am Med Inform Assoc. 2011; 18(3):282–91. [PubMed: 21486887]
28. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc. 2011; 18(2):181–6. [PubMed: 21233086]
29. Linder, JA.; Schnipper, JL.; Middleton, B. [Accessed July 2012.] Method of electronic health record documentation and quality of primary care. J Am Med Informat Assoc. 2012. Available at <http://jamia.bmj.com/content/early/2012/05/19/amiainl-2011-000788.abstract>
30. Parsons, A.; McCullough, C.; Wang, J.; Shih, S. [Accessed July 2012.] Validity of electronic health record-derived quality measurement for performance monitoring. J Am Med Informat Assoc: JAMIA. 2012. Available at <http://jamia.bmj.com/content/early/2012/02/08/amiainl-2011-000557.long>
31. Hoff T. Deskillling and adaptation among primary care physicians using two work innovations. Health Care Manage Rev. 2011; 36(4):338–48. [PubMed: 21685794]
32. Zheng K, Hanauer DA, Padman R, Johnson MP, Hussain AA, Ye W, et al. Handling anticipated exceptions in clinical care: investigating clinician use of “exit strategies” in an electronic health records system. J Am Med Inform Assoc. 2011; 18(6):883–9. [PubMed: 21676941]
33. Freudenheim, M. Digitizing Health Records Before It Was Cool. New York Times; Jan 14. 2012 Available at http://www.nytimes.com/2012/01/15/business/epic-systems-digitizing-health-records-before-it-was-cool.html?_r=1&pagewanted=all
34. Lanham HJ, Leykum LK, McDaniel RR. Same organization, same electronic health records (EHRs) system, different use: exploring the linkage between practice member communication patterns and EHR use patterns in an ambulatory care setting. J Am Med Informat Assoc. 2012; 19(3):382–391.

35. Institute of Medicine. [Accessed June 2012.] Informatics Needs and Challenges in Cancer Research: A Workshop. 2012. Available at <http://www.iom.edu/Activities/Disease/NCPF/2012-FEB-27.aspx>
36. George, A. [Accessed July 2012.] Komatsoulis. Program Announcement — caBIG. 2012. Available at https://cabig.nci.nih.gov/program_announcement
37. Varmus, Harold. [Accessed July 2012.] Announcement: New National Cancer Informatics Program (NCIP) — caBIG. 2012. Available at https://cabig.nci.nih.gov/ncip_announcement/
38. President's Council of Advisors on Science and Technology. Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward. Washington, DC: Executive Office of the President; 2010. Available at <http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf>
39. Chen D, Doumeingts G, Vernadat F. Architectures for enterprise integration and interoperability: Past, present and future. *Computers in Industry*. 2008; 59(7):647–59.
40. Institute of Medicine (US) Committee on Comparative Effective Research. Initial national priorities for comparative effectiveness research. National Academies Press; 2009.
41. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med*. 2009; 151(3):203–5. [PubMed: 19567618]
42. [Accessed July 2012.] Healthcare Information and Management Systems Society - HIMSS. Available at <http://www.himss.org/ASP/index.asp>
43. Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. *New England Journal of Medicine*. 2010; 363(6):501–4. [PubMed: 20647183]
44. Foundation for Health Services Research. Health Outcomes Research: A Primer. Washington, DC: Foundation for Health Services Research; 1994. p. 10 Available at <http://www.academyhealth.org/files/publications/healthoutcomes.pdf>