Published in final edited form as:

Med Care. 2013 August ; 51(8 0 3): S38–S44. doi:10.1097/MLR.0b013e31829b1de1.

Knowledge Management and Informatics Considerations for Comparative Effectiveness Research: A Case-driven Exploration

Peter J. Embi, MD, MS¹, Courtney Hebert, MD¹, Gayle Gordillo, MD², Kelly Kelleher, MD³, and Philip R.O. Payne, Ph.D¹

¹The Ohio State University, Department of Biomedical Informatics, 333 W. 10th Avenue, Columbus Ohio, 43210

²The Ohio State University, Department of Plastic Surgery, 915 Olentangy River Road, Columbus OH, 43212

³Nationwide Children's Hospital, 700 Children's Drive, Columbus Ohio, 43205

Abstract

Background—As clinical data are increasingly collected and stored electronically, their potential use for comparative effectiveness research (CER) grows. Despite this promise, challenges face those wishing to leverage such data. In this paper we aim to enumerate some of the knowledge management and informatics issues common to such data re-use.

Design—After reviewing the current state of knowledge regarding biomedical informatics challenges and best practices related to CER, we then present two research projects at our institution. We analyze these and highlight several common themes and challenges related to the conduct of CER studies. Finally, we represent these emergent themes.

Results—The informatics challenges commonly encountered by those conducting CER studies include issues related to data information and knowledge management (e.g. data re-use, data preparation) as well as those related to people and organizational issues (e.g. socio-technical factors and organizational factors). Examples of these are described in further detail and a formal framework for describing these findings is presented.

Conclusion—Significant challenges face researchers attempting to use often diverse and heterogeneous datasets for CER. These challenges must be understood in order to be dealt with successfully and can often be overcome with the appropriate use of informatics best practices. Many research and policy questions remain to be answered in order to realize the full potential of the increasingly electronic clinical data available for such research.

Keywords

Knowledge Management; comparative effectiveness research; medical informatics

Corresponding Author: Peter J. Embi, MD, MS, The Ohio State University, 3190 Graves Hall, 333 W. 10th Ave, Columbus, OH 43210, Telephone: 614-292-4778, Fax: 614-688-6600, peter.embi@osumc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The ability to combine and analyze heterogeneous and multi-dimensional datasets is central to the design and conduct of CER studies (1). The generation of such datasets usually involves aggregating diverse but related data types, including but not limited to: 1) clinical observations and outcomes; 2) operational and fiscal indicators; 3) bio-molecular markers; and 4) patient-reported variables. The ability to reproducibly and systematically identify, integrate, and deliver such datasets involves biomedical informatics technologies and approaches such as information retrieval, text mining, temporal reasoning and semantic normalization operations. Unfortunately, the knowledge and expertise needed to understand, select, and apply such methods is often distributed across a variety of professional domains, literature corpora, and prevailing best-practice collections. While relevant to many types of research, these challenges are further amplified in the context of CER, given the need to identify and utilize often broader classes of data than have been the focus of traditional research paradigms.

Given these gaps in knowledge and practice, this report has the following objectives: 1) to review the major challenges associated with the collection and aggregation of heterogeneous and multi-dimensional datasets 2) to propose a conceptual model for the design and conduct of CER; and 3) to introduce a set of open research questions and policy priorities for the adoption of relevant information practices.

2. Background

To address specific challenges in managing heterogeneous datasets for CER, one must first have an understanding of the informatics theories, methods and frameworks that are available for implementation in such an application domain. These include methods to handle data discovery, extraction and integration. In the following section we will outline some of these methods and discuss how they can be translated for use in CER.

2.1 Data discovery and extraction

The science of managing biomedical data has advanced significantly in recent decades. Trends or patterns in large datasets can increasingly be exposed using data-mining technologies and methods that have been adopted for wide ranging uses in biomedicine. (2, 3) For CER, such technologies and approaches are particularly useful for hypotheses generation, study design and feasibility assessment, and cohort identification activities like participant recruitment.

Data from electronic health records (EHR), laboratory data and billing data are generally collected and stored in transactional databases intended for patient care and administrative use, not research. Data warehousing provides a method for health care organizations to combine clinical, laboratory, and administrative data in a manner that can then be queried for secondary purposes, such as research. (4) A data warehouse should be subject oriented, time dependent, nonvolatile and well integrated. (5) Many organizations have adopted data warehouses allowing for sophisticated data driven research. Some have made it possible for researchers to interact directly with de-identified data available to researchers, enabling their data to be leveraged for efficient hypothesis generation, study design and feasibility assessments, and certain research studies. (6)

Another approach to data integration and aggregation is the utilization of reusable computational pipelines where users can gather information from multiple institutions and databases, and share constituent data between and among sites. For example, techniques such as service-oriented architectures (SOA) have been used to create community accessible

data pipelines. One SOA platform in the research domain is the National Cancer Institute's Cancer Biomedical Informatics Grid (7) program, which uses a SOA infrastructure named caGrid to provide a common, collaborative, extensible platform for cancer researchers to allow for data interchange and analysis (8).

More recently, the computational and informatics community has begun to explore the use of highly scalable and elastic “cloud computing platforms for data sharing and analysis, using architectural approaches commonly referred to as “software as a service or “platform as a service. In both of these scenarios, data exchange, storage, and analysis are achieved through the use of remotely managed and standardized computing resources, accessible via common Internet technologies and facilitated through high-speed networking capabilities. (9) Even as such technological approaches to data discovery are maturing, relevant data must be stored in a manner that can be readily queried and analyzed.

Clinical documentation may contain the data needed for research, but those data are often stored in an unstructured format. While trained research personnel have traditionally been employed to extract data from clinical documents for research, automated extraction is highly desired for many anticipated CER activities. Improvements in natural language processing (NLP) of narrative documents are ongoing and hold great promise, but current NLP approaches often still require clinician input and have yet to gain widespread adoption (10-14).

Regardless of the effectiveness of such approaches, it remains the case that the quality of research is dependent on the quality of the data that is captured in the EHR or other databases. Kahn and colleagues recently proposed a conceptual model for data quality assessment which suggests specific methods for ensuring quality data from EHR sources (15). These methods can be implemented on the individual level or automated to work across a database to assess the quality of data elements.

Improving systematic data capture by clinicians and directly by patients for data re-use, without disrupting work flows remains an important national goal (16). In addition, researchers are beginning to use increasingly ubiquitous mobile devices, telemedicine, and personal monitoring technologies to collect patient-derived data. (17) A detailed review of these methods is beyond the scope of this paper; however, it is reasonable to expect that these technologies will continue to improve and enhance structured data for CER in the near future.

2.2 Data integration

Problems can arise when trying to combine data with different storage structures or ontologies, or data from different and incompatible systems. Semantic normalization is the process of normalizing terminology across different datasets in order to be able to combine datasets while maintaining the appropriate meaning of the data elements. This process has been used to integrate structured and unstructured data (18) as well as to combine large genomic datasets with phenotypic data (19). The key to integration of diverse data sources is to have a clear understanding of the type and structure of data in each database. The use of relational databases with data models specifically designed for EHR data as well as the use of multi-site informatics platforms can facilitate data integration for researchers (20, 21).

The timing of clinical events also needs to be taken into account when combining datasets. Although laboratory and encounter level data come with a time stamp, text based data handle time more subjectively. Determining the order of symptoms in a timeline is a major challenge when trying to integrate lab and clinical data. Temporal reasoning techniques have been developed to help deal with these challenges (22).

In the sections that follow, we start by exploring two example cases. These cases illustrate common themes that are encountered in CER, including issues of: (1) data discovery (2) data warehousing (3) temporal reasoning and (4) integration of datasets. Each case also illustrates unique challenges based on the specific type of data collected.

3. Methods and Approach

For this effort, we pursued a multi-step process (Figure 1). We engaged two distinct groups of subject-matter experts (SMEs) from relevant research teams at Ohio State University Wexner Medical Center (OSUWMC) and Nationwide Children's Hospital (NCH). Three authors (PJE, CH, PROP) conducted a series of semi-structured interviews and analyzed notes from those sessions. Based on that analysis, we generated summaries of the cases, identified key challenges as well as emergent themes. We then used these examples and findings to anchor a discussion of future directions for addressing the informatics issues related to design and conduct of CER.

4. Example cases

In the following sub-sections, we summarize each of the two example research cases, discuss issues related to data discovery and data integration, and address any unique challenges that were encountered.

4.1 - Case 1: Standard versus specialized wound care for patients with diabetes

The objective of this study was to determine whether patients with diabetic foot ulcers have a lower amputation rate when treated by a wound care specialist versus standard care from providers without specific wound care training. The primary objective was to compare amputation rates between these two groups, stratified by a validated ulcer severity score, the Wagner score.

4.1.1 Data discovery and extraction—The data needed for the study were collected from two main sources. For inpatients receiving standard care the data were taken from the OSUWMC Information Warehouse (IW) and enterprise data warehouse that is used for both operational and research purposes. For the patients receiving specialized care, the data originated in outpatient clinics. The National Healing Corporation manages these data, and then sends the data to the OSUWMC IW.

4.1.2 Data integration—Since the data for this study were extracted from two different datasets for the two comparison groups, one outpatient and one inpatient, it was a challenge to consistently define variables between the two groups. For example, procedures were classified using International Classification of Diseases (ICD-9) procedure codes for the standard care group and Current Procedural Terminology codes (CPT) for the outpatient specialty care group. The research team created an algorithm to map the different coding systems. In addition, because of differences between the data sources, there were few ways to comparably define comorbidities for the two groups. The researchers decided to use laboratory values to define comorbidities because these were some of the few data elements that were consistently recorded between the two groups. This practice can introduce error into a study. Laboratory values may be only temporarily abnormal and abnormal values are not always reflective of comorbid conditions especially in the absence of clinical interpretation.

4.1.3 Unique challenges—Beyond the more commonly recognized issues, several unique challenges were encountered. First was the issue of data completeness and consistency. For instance, once the data were collected it was found that Wagner scores

were only available on specialty patients. The result was that standard-care patients had to have Wagner scores retrospectively calculated, likely introducing bias when compared to the prospectively calculated specialty care patients.

A major challenge involved the primary outcome of the study. The initial primary outcome of the study was wound healing. However determining the instant when a wound is healed over the course of several encounters was not possible with the current documentation. Although some notes included photo documentation, many did not. In the end amputation was chosen as a surrogate outcome because it could be easily extracted from available data. Data availability leading to changes in the study's primary outcome is just one consequence of the current state of informatics for CER.

4.2 - Case 2: Comparing primary care interventions for teenagers at risk for depression

The second study was a study of waiting room tablet-based behavioral health screening of teenagers at risk for depression and drug abuse. The study compared immediate physician notification of screening results versus delayed notification. It examined the efficacy of screening, the success of referrals and provision of mental health care. Ultimately the goal was to evaluate if these interventions would increase the number of at-risk youths seeking appropriate mental health care.

4.2.1 Data discovery and extraction—The data used for this study came from multiple sources. These sources included: (1) paper-based patient visit surveys (2) billing data (3) nurse telephone logs (4) transcribed telephone calls (5) school records (6) insurance records (7) behavioral health clinic documentation. Combining these different sources of data posed significant challenges in assessing outcomes. For instance, youth reports of depression or drug treatment sessions significantly underestimated those based on billing data. Similarly, parents underreported services received at primary care or school offices. Because of the diversity of sources and the often-conflicting data, the research team needed to make determinations of data provenance.

4.2.2 Integration—Though there were many electronic data sources for this study, creating a single database was complicated by several factors. No single dataset contained a common identifier; thus, name and birth date combinations were the only common linkage and this was inadequate in several cases. Some children used different names depending on parent residence. Further, the diverse types of data from disparate systems including schools, medical and law enforcement organizations made full integration impossible.

4.2.3 Unique Challenges—Obtaining data from multiple sources proved to be a challenge. Insurance companies, schools, law enforcement and child welfare offices were uncomfortable with requests for child-specific data, especially data that focused on older records. They did not feel comfortable fulfilling such requests and often delayed or did not respond. A number of insurance companies and some schools simply refused to include data from their sites.

Even when data were available, unique pediatric-specific issues arose. Because the study included patients over a range of ages and developmental milestones, standardizing risk factors and developmental milestones over the study population was complicated. Due to the rapid development of teenagers and young children temporal adjustments had to be made when examining laboratory data, weight and height, and other milestones.

5. Results: Emergent Themes

Based upon the lessons learned from the available literature and example cases described above, we find that a conceptual model of common factors influencing the design and conduct of CER emerges. This conceptual model includes four major factors. Two are intrinsic to the data (data quality, data preparation) and two that are extrinsic (socio-technical factors and organizational factors). The major elements are illustrated in Figure 2 and described in detail below:

5.1 Thematic Category 1: Data quality

There are many data quality issues that arise when re-using data for research as outlined in Kahn's conceptual model of data quality (15). The following factors are those that are most commonly encountered in CER.

5.1.1 Data availability—EHRs were not designed to gather data for research but rather to enable healthcare processes. Moreover, healthcare providers are typically neither trained to record data for research purposes nor are they encouraged to do so. (23) These factors often result in missing data. In case 1, the major effect of this was missing Wagner scores. Inpatient physicians were not aware of the importance of this scoring system when evaluating patients with diabetic foot ulcers. In case 2 the researchers sought to include school and juvenile detention data. While such community-derived data are often essential to certain research efforts, regulatory challenges and unstructured formatting of such data made collection and analysis especially difficult.

5.1.2 Lack of shared semantics—This is a common problem that is illustrated in both our cases. In case 1, the researchers came up with their own protocol for translating ICD-9 codes into CPT codes so that the two groups of patients could be accurately compared. In case 2 the group was unable to reliably link patients from one database to another because of a lack of a unique identifier.

5.1.3 Determination of data provenance—A common problem with using electronic records for CER is defining variables that have more than one possible operational definition. For example, does a patient have chronic renal disease because he has an billing code (ICD-9) recorded for renal disease, he has seen a renal specialist, he has an elevated creatinine on one or more lab draws or “renal disease is written in free text in his progress note? In case 1, the researchers chose to use laboratory values to determine comorbidities. This is attractive in that it seems to be an objective measure, however laboratory values can change over time, can be only temporarily abnormal, and deciding how to classify patients with missing laboratory values can be difficult. In case 2, data provenance was an issue when determining which document represented the “truth regarding whether a teenager saw a mental health professional. The research team decided to use an agreed upon protocol in order to standardize their data collection; however, more formal data provenance structures and reusable pipelines can help standardize conflicting data sources.

5.2 Thematic Category 2: Data preparation: Temporal relationships and Data synthesis

When data is re-used, it is often necessary to process the data in a way that maximizes its applicability to the study. One of the major processing steps involves dealing with temporal relationships. In case 1, the team was unable to use wound healing as a primary outcome because determining an exact time of wound healing was not possible. Amputation, a more objective though less satisfying outcome was chosen instead.

In case 2, temporal issues were encountered due to the need to account for aging of their patient population over time. Developmental milestones or laboratory values that might be normal for a patient at one point in the study could be pathological at another.

5.4 Thematic Category 3: Socio-technical factors

These factors were mostly encountered in case 2 where regulatory frameworks were obstructive and prevented some participants from contributing data to the study. Because the most difficult data to obtain for this study came from juvenile law enforcement, the children who could have benefitted and contributed the most were not included in the study. Although standards must be in place to protect patients' privacy, in order to be able to leverage this data for CER, institutional and regulatory entities need to be receptive to calls for quality improvement in data management and easing regulatory frameworks that impede data sharing.

5.3 Thematic Category 4: Organizational factors

Organizational factors include availability of trained informatics personnel involved in database management as well as institutional support. In case 2, the data were being gathered and processed by untrained personnel. This can introduce error as well and inconsistencies in the data. A well functioning data warehouse or database management group is essential for CER.

6. Discussion

As the above case examples illustrate, there remain many knowledge management and informatics challenges to the systematic, routine and successful leveraging of heterogeneous clinical data for CER. Those conducting or supporting the conduct of CER would benefit from a framework upon which to approach the often challenging informatics issues commonly encountered when engaging in such research involving clinical data re-use.

Even with trained personnel and a supportive institution, the current state of clinical data is often insufficient or inadequate to address the information needs of researchers. There are numerous open informatics research and policy related questions that must be addressed in order to fully realize the potential of what we anticipate and expect CER endeavors to yield in terms of better, safer, and more cost-effective healthcare delivery. (24)

6.1 Open informatics research questions

It is clear that informatics research in this domain must include those issues related to data discovery and integration. Questions remain as to the best way to store and structure data in order to be able to use it for research purposes, especially in data re-use scenarios. Vital to this question is how best to capture data from key members of the healthcare team in order to have research quality data available. This may involve improving the ease of data entry for clinicians and/or improving methods for collecting high-quality data directly from patients. Developing new methods for integrating data from other non-traditional healthcare sources also holds promise and must be pursued, particularly for the conduct of certain types of specialty studies.

Similarly, improved and standardized protocols need to be established to determine hierarchical and semantic relationships between different data sources and types. Standards for dealing with missing data also need to be developed and incorporated into practice. In addition, fundamental methodological innovation is needed in the collection and integrative analysis of heterogeneous and multidimensional data, especially as they relate to knowledge engineering, temporal reasoning, data synthesis (including natural language processing),

knowledge discovery in databases, and the visualization of complex and information dense constructs.

6.2 Open policy issues

Data for CER are currently stored in various locations and with various semantic structures. As our case examples illustrate, significant effort is often needed to extract and combine data from multiple sources. Too often these efforts are made more difficult by prevailing policies rather than by technical limitations. Even as technical issues must be addressed to facilitate such endeavors, policy and regulatory issues are critical to realizing the benefits of such data, and many can be readily addressed today.

For one, policies that promote and enforce adherence to and use of standards across institutions are essential and must be addressed. Within individual healthcare organizations, investments to improve analytical and business intelligence capacities should be viewed for their ability to simultaneously enable both operational and research activities alongside one another, for the benefit of both. Along similar lines, interpretations of existing policies and regulations regarding data re-use for research that often impede research productivity should be re-thought to safely and securely enable such work (25).

On a broader scale, efforts to make data sharing more standardized and reproducible across institutions would benefit greatly from adjustments and reconciliation of regulatory frameworks that now often impede such sharing if it is for anything other than clinical care purposes. At the individual level, policies and incentive structures that promote and foster the involvement of clinicians and other members of the healthcare team in the systematic collection of data and engagement of participants in research are also sorely needed (23).

Finally, in order to accomplish these goals, we will need to have professionals who are trained in informatics and clinical research. Currently there is a shortage of such highly-trained professionals and a paucity of mechanisms to support or enable critically needed training programs in this domain.

6.3 Limitations

This paper has several limitations. First, our description of the current knowledge base is not intended to be a systematic review of the literature but rather a targeted review aimed at informing current issues associated with CER. We did not focus on the foundational work in the computational and information sciences, as this is a paper aimed principally at those conducting and supporting CER. Also, while we believe that most of these problems are common in CER, some of the problems discussed may be reflective of issues unique to our institution. In addition, although our conceptual model presents four individual thematic categories, in real-use there is significant interplay between these factors, adding to the complexity of CER. Finally, it should be noted that the limited number of exemplary cases and few SMEs involved means that there may be other factors we did not capture via this approach.

7. Conclusion

The conduct of CER can benefit greatly from leveraging the increasingly electronic sources of clinical data available. However, there remain many informatics challenges to overcome if we are to realize the opportunities facing the healthcare and biomedical research communities. Those engaged in such endeavors should consider the issues illustrated above as they plan and pursue CER studies. This proposed framework should help to inform the considerable work that remains to be done at the intersection of biomedical informatics, data management and CER.

Acknowledgments

Funding: This work was funded by a grant from the EDM forum. Case 1 mentioned in the paper was funded by the CCTS (NIH UL1RR025755) and Case 2 was funded by a grant from the National Institute on Drug abuse (5R01DA018943-04)

References

1. Initial National Priorities for Comparative Effectiveness Research. The National Academies Press; 2009.
2. Woeltje KF, Lautenbach E. Informatics and epidemiology in infection control. *Infectious disease clinics of North America*. 2011; 25:261–270. [PubMed: 21316004]
3. Marks L, Power E. Using technology to address recruitment issues in the clinical trial process. *Trends in biotechnology*. 2002; 20:105–109. [PubMed: 11841861]
4. Hu H, Correll M, Kvecher L, et al. DW4TR: A Data Warehouse for Translational Research. *Journal of biomedical informatics*. 2011; 44:1004–1019. [PubMed: 21872681]
5. Embi PJ, Kaufman SE, Payne PR. Biomedical informatics and outcomes research: enabling knowledge-driven health care. *Circulation*. 2009; 120:2393–2399. [PubMed: 19996023]
6. Erdal BS, Liu J, Ding J, et al. A Database De-identification Framework to Enable Direct Queries on Medical Data for Secondary Use. *Methods of information in medicine*. 2012; 51
7. Institute NC. [Accessed December 26, 2012] Digital Capabilities to Accelerate Research. Available at: <https://cabig.nci.nih.gov>
8. Saltz J, Oster S, Hastings S, et al. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics*. 2006; 22:1910–1916. [PubMed: 16766552]
9. Rosenthal A, Mork P, Li MH, et al. Cloud computing: A new business paradigm for biomedical information sharing. *Journal of biomedical informatics*. 2010; 43:342–353. [PubMed: 19715773]
10. Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. *Journal of biomedical informatics*. 2006; 39:589–599. [PubMed: 16359928]
11. Chen, ES.; Hripcsak, G.; Friedman, C. Disseminating natural language processed clinical narratives. *AMIA Annual Symposium proceedings/AMIA Symposium* AMIA Symposium; 2006; p. 126-130.
12. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction *Journal of the American Medical Informatics Association*. *Journal of the American Medical Informatics Association: JAMIA*. 2011; 18:544–551. [PubMed: 21846786]
13. Webb KH. Natural language processing and electronic medical records. *JAMA: the journal of the American Medical Association*. 2011; 306:2325. author reply 2325-2326. [PubMed: 22147375]
14. Liu, F.; Weng, C.; Yu, H. Natural Language Processing Electronic Health Records Clinical Research. In: Richesson, RL.; Andrews, JE., editors. *Clinical Research Informatics*. Springer; 2012. p. 293-310.
15. Kahn MG, Raebel MA, Glanz JM, et al. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Medical care*. 2012; (50 Suppl):S21–29. [PubMed: 22692254]
16. Cusack CM, Hripcsak G, Bloomrosen M, et al. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. *Journal of the American Medical Informatics Association: JAMIA*. 2013; 20:134–140. [PubMed: 22962195]
17. Hampton T. Recent advances in mobile technology benefit global health, research, and care. *JAMA: the journal of the American Medical Association*. 2012; 307:2013–2014. [PubMed: 22665083]
18. Mirhaji P, Zhu M, Vagnoni M, et al. Ontology driven integration platform for clinical and translational research. *BMC bioinformatics*. 2009; 10(2):S2. [PubMed: 19208190]
19. Goh CS, Gianoulis TA, Liu Y, et al. Integration of curated databases to identify genotype-phenotype associations. *BMC genomics*. 2006; 7:257. [PubMed: 17038185]

20. Sittig DF, Hazlehurst BL, Brown J, et al. A survey of informatics platforms that enable distributed comparative effectiveness research using multi-institutional heterogeneous clinical data. *Medical care*. 2012; (50 Suppl):S49–59. [PubMed: 22692259]
21. Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. *Medical care*. 2012; (50 Suppl):S60–67. [PubMed: 22692260]
22. Zhou L, Hripcsak G. Temporal reasoning with medical data--a review with emphasis on medical natural language processing. *Journal of biomedical informatics*. 2007; 40:183–202. [PubMed: 17317332]
23. Embi PJ, Tsevat J. Commentary: the relative research unit: providing incentives for clinician participation in research activities. *Academic medicine: journal of the Association of American Medical Colleges*. 2012; 87:11–14. [PubMed: 22201633]
24. Payne PR, Embi PJ, Niland J. Foundational biomedical informatics research in the clinical and translational science era: a call to action. *Journal of the American Medical Informatics Association: JAMIA*. 2010; 17:615–616. [PubMed: 20962120]
25. Ness RB. Influence of the HIPAA Privacy Rule on health research. *JAMA: the journal of the American Medical Association*. 2007; 298:2164–2170. [PubMed: 18000200]

Figure 1. Overview of methodology used to assess common data, information, and knowledge management factors encountered during the design and conduct of CER studies

Figure 2. Conceptual model of data, information, and knowledge management factors influencing and/or predisposing the design and conduct of CER studies