Opportunities and Challenges for Comparative Effectiveness Research (CER) With Electronic Clinical Data

A Perspective From the EDM Forum

Erin Holve, PhD, MPH, MPP, Courtney Segal, BA, and Marianne Hamilton Lopez, MPA

Introduction: The Electronic Data Methods (EDM) Forum brings together perspectives from the Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies (PROSPECT) studies, the Scalable Distributed Research Networks, and the Enhanced Registries projects. This paper discusses challenges faced by the research teams as part of their efforts to develop electronic clinical data (ECD) infrastructure to support comparative effectiveness research (CER). The findings reflect a set of opportunities for transdisciplinary learning, and will ideally enhance the transparency and generalizability of CER using ECD.

Methods: Findings are based on 6 exploratory site visits conducted under naturalistic inquiry in the spring of 2011. Themes, challenges, and innovations were identified in the visit summaries through coding, keyword searches, and review for complex concepts.

Results: The identified overarching challenges and emerging opportunities include: the substantial level of effort to establish and sustain data sharing partnerships; the importance of understanding the strengths and limitations of clinical informatics tools, platforms, and models that have emerged to enable research with ECD; the need for rigorous methods to assess data validity, quality, and context for multisite studies; and, emerging opportunities to achieve meaningful patient and consumer engagement and work collaboratively with multidisciplinary teams.

Discussion: The new infrastructure must evolve to serve a diverse set of potential users and must scale to address a range of CER or patient-centered outcomes research (PCOR) questions. To achieve this aim—to improve the quality, transparency, and reproducibility of CER and PCOR-a high level of collaboration and support is necessary to foster partnership and best practices as part of the EDM Forum.

Key Words: comparative effectiveness research, learning health care system, scalability, sustainability, stakeholder, transdisciplinary

(*Med Care* 2012;50: S11–S18)

From the EDM Forum (Project) the Research and Education in Health Services Research Division at AcademyHealth, Washington, DC.

Supported by AHRQ Grant U13 HS19564-01 (Electronic Data Methods Forum for Comparative Effectiveness Research).

The authors declare no conflict of interest.

Reprints: Erin Holve, PhD, MPH, MPP, AcademyHealth, 1150 17th Street NW, Suite 600, Washington, DC 20036. E-mail: erin.holve@academyhealth.org.

Copyright © 2012 by Lippincott Williams & Wilkins

ISSN: 0025-7079/12/5007-0S11

This paper discusses cross-cutting challenges and opportunities for research projects participating in the Electronic Data Methods (EDM) Forum. The EDM Forum is a 3-year grant from the Agency for Healthcare Research and Quality to facilitate learning and foster collaboration across a set of 11 comparative effectiveness research (CER) projects. Further information and details on the EDM Forum and collaborating projects is included in a separate report by Holve and colleagues in this special supplement. Findings in this paper are based on a series of 6 site visits with the Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies (PROSPECT) studies, the Scalable Distributed Research Networks (DRN) for CER, and the Enhanced Registries for Quality Improvement (QI) and CER projects that were conducted to identify common issues when using electronic clinical data (ECD) for CER. Each of the themes identified in the paper is intended to provide a summary of early efforts across the teams, and serve as a reference to reflect upon over the course of the grants. These lessons will ideally enable the scientific community to learn from the projects' experiences and innovative efforts, and also increase the transparency and generalizability of the science. Appendix provides a list of acronyms and terminology to define the breadth of approaches and innovative strategies discussed throughout this paper.

To engage a discussion around CER using ECD it is important to note that as a multidisciplinary area of inquiry, CER arguably requires a *transdisciplinary*[1] approach to evidence generation that integrates the perspectives of multiple disciplines such as outcomes research, biomedical informatics, statistics, and specialized clinical perspectives, among others. By the nature of disciplines each tends towards particular approaches and methods, and uses specialized vocabulary. However, building infrastructure to conduct CER and patient-centered outcomes research (PCOR) requires that professionals bridge disciplinary silos to foster rapid-cycle innovation. Rapid-cycle innovation is a method by which improvement tools and techniques can be designed and tested, with the goal of eliminating process steps that do not hold value.[2] Rapid-cycle innovation requires that timely connections are made between innovators, and that useful collaborative resources are built in the service of shared scientific goals. The findings from the site visits reflect timely and practical

efforts by the PROSPECT, DRN, and Enhanced Registry projects to achieve a level of dialog and collaboration that will advance the science of CER using ECD.

METHODS

The findings in this paper are based on 6 site visits conducted by EDM Forum staff at research, medical, and academic centers in the spring of 2011. The research projects included in the site visits are all participants in the EDM Forum and were selected to represent a variety of characteristics across the research networks (eg, geographic scope; age and maturity of the network). Because of the size of the networks, coordinating sites where the key investigators could be interviewed served as the unit of analysis. These visits were highly exploratory, and focused on challenges and innovations across the projects, as well as areas for collaboration and shared learning. To ensure confidentiality, individual projects and the characteristics of specific sites have been suppressed, and all interviews have been stripped of identifiers. Information about the 11 research projects participating in the EDM Forum can be found at http://www.edm-forum.org.

Data Collection

AcademyHealth received an Institutional Review Board (IRB) exemption for this project from the Western Institutional Review Board, based on an interview and data collection process designed to ensure the confidentiality of all study participants and information shared during the visits. To prepare for each of the visits, the EDM Forum staff conducted a thorough review of all the 11 teams' research plans and the projects' websites (if applicable), as well as attended presentations given by each of the teams that provided an overview of their respective projects. On the basis of this preparatory work, the EDM Forum staff developed a set of general questions for all of the sites, and specific questions for each project (see Table 1 for examples).

The site visits were conducted under a naturalistic inquiry and emergent design to allow for flexibility in the structure of the visits as they proceeded, acknowledging that the discussions would be inductive in nature.[3] Interviews and discussions were conducted over a full day at the research project's coordinating site. Each site visit began with a presentation by EDM Forum staff to reflect our understanding of the project and clarify aspects of each project that may have been misinterpreted or that we failed to consider. At most sites, project staff presented updates related to key questions of interest for the EDM Forum (eg, project management, technical approaches to validating data, security protocols and measures, community engagement efforts, etc.). All visits included ≥ 10 investigators (representing various facets of the project) over the course of each visit. Conversations were not recorded, but multiple note takers summarized comments throughout the day, and in many cases, presentation slides were provided by the sites to augment note-taking.

Coding, Analysis, and Reporting

On the basis of the first (pilot) site visit, staff developed an initial coding scheme and framework for note-taking. For the pilot and all subsequent visits, confidential site visit summaries were prepared by organizing findings based on the initial coding scheme. As new areas of emphasis emerged in subsequent visits, these topics generated "subcodes" that were added to the coding scheme. Each of the summaries was then shared with project leads to assess the accuracy of the summaries and provide comments on categorization of findings. At the conclusion of all of the visits and subsequent to receiving feedback from project leads, the visit summaries were compiled and keyword searches (as well as a manual review for complex concepts) were conducted using a code list based upon all codes generated throughout the visits. This approach helped to ensure consistency in coding and provided a level of confidence that we had reached saturation of major themes.

RESULTS

Findings from the site visits suggest that efforts to build infrastructure to support the use of ECD for CER face 4 overarching challenges, and an important set of emerging opportunities. These 4 issues include: (1) the substantial effort and resources the projects have employed to establish and sustain data sharing partnerships; (2) a range of clinical informatics tools, platforms, and models have been developed to enable research with ECD, *but* there is a need to understand the strengths and limitations of each; (3) the sites see a need for rigorous methods to assess data validity, quality, and context for multisite studies given the absence of a gold standard for evaluating ECD; (4) there are new opportunities to achieve meaningful patient and consumer engagement and work collaboratively with multidisciplinary teams.

The Projects are Employing Substantial Effort and Resources to Establish and Sustain Data Sharing Partnerships

To conduct rigorous research, all of the project teams mentioned that building trust among data sharing partners is paramount. Developing governance structures that support these relationships and allow "data to flow" requires an ongoing and time intensive effort, and explicit infrastructure support for project management. In particular, project staff emphasized the importance of relationships in facilitating IRB and Privacy Board approval across multiple institutions.

Several of the investigators commented that the explicit financial support for "infrastructure building" in the American Recovery and Reinvestment Act awards for the PROSPECT, DRN, and Enhanced Registry projects should be acknowledged. CER requires a broad range of technical expertise and each of the networks has relatively large project teams, with 20–40 staff including clinical investigators, data programmers, informaticians, biostatisticians, epidemiologists, clinicians, and other team members contributing various amounts of time. In this

 © 2012 Lippincott Williams & Wilkins

TABLE 1. General Areas of Inquiry by Theme

Before each site visit questions from the following list (including some specific to the site) were sent to the research teams to give them a general sense of our interest. EDM Forum staff referred to the list of questions throughout the day (when relevant to the conversation).
Due to the exploratory nature of the site visits, additional questions were asked to understand the teams' approaches to building infrastructure and undertaking innovative approaches for analytic methods, clinical informatics, and data governance.
Data Governance
What types of security protocols have you put in place for the study, particularly for the data collection tools?
How does your team (including IT) work with the IRB and/or Privacy Boards at the institutions participating in the project?
How will data governance be handled for future studies using the data and the infrastructure?
How do you define the distinction between QI and research, and how will the determination that a project is "QI" or "research" be reached? Handled?
What are the issues you are facing in the process of obtaining consent for QI or research purposes?
What types of policies and governance structures are in place to ensure the registry/network/infrastructure is sustainable across such a large network?
How is the use of EHR data for the project's research objectives communicated to the patient at the point of care?
How will consent be obtained when the infrastructure and the data collected for this project is open for future CER activities beyond the studies conducted during this three-year grant?
Informatics
Please talk about why you elected to federate or distribute queries across sites (if applicable).
What were your criteria for selecting the informatics platform you are using (including the assessing performance of the tool)? Were there other informatics platforms for research that you considered?
Given the range of data available in your network and/or research data warehouse, how will you standardize vocabulary across sites and sources?
What types of metadata are you collecting as part of the project?
How will you ascertain researchers' needs for accessing the data? What is your vision for how researchers will use the tool outside of this study (if applicable)?
Please talk about the relationship between the EHR Vendors used at the project sites and their awareness of research activities.
If some data partners have limited access to programming resources, what challenges do you foresee in bringing these partners into your project/network?
Analytic Methods
How will you validate the data (especially the longitudinal measurements) collected from PRO against information collected during clinical care in the EHR?
Do you foresee any analytic challenges when using de-identified data for analysis?
Please talk about the innovative methods to incorporate PRO information that are typically captured on paper-based formed and not routinely entered into EHRs. How will you validate the new source of data against the currently available paper-based data?
Please talk about how you will monitor data quality of the near-real time data collection. Please talk about the validation measures at the single-site and multi-site levels.
How will you assess the data quality of health care delivery factors and patient-reported outcomes implemented into the EHR?
What impact, if any, will your choice and approach with the data model have on your analytic plan?
Community Engagement
How did the collaboration with the participating institutions and/or data sharing partners developed? Over what time horizon?
What does the team think are the true requirements are (of the research project/systems, etc.) in order to meaningfully engage and partner with the community?
How do you plan to engage stakeholders in building this effort on an ongoing basis? How do you define this group?
Are there strong social networks or common sources of information that providers and patients turn to for information on the priority condition(s) your CER studies are addressing? Do you feel this is an issue you will need to take into account in your analysis, particularly with the PRO data?
Learning Healthcare System & Quality Improvement
From the quality improvement and research perspective, what is your vision for the way in which this project can contribute new knowledge and information to improve health and health care? In other words, what does this new infrastructure gets you that current research activities do not?
Is the importance of research transparent across the non-academic partners involved in the project? How do QI or research findings get back to the centers?
How do you plan to utilize the analyses from the CER studies in quality improvement initiatives for healthcare delivery?
Other
Are there steps you are taking to ensure the sustainability of the infrastructure/registry/network?
What types of training needs do you foresee for the research community to enable best use of the types of data resources you are developing?
Are there any unique innovations and/or challenges that emerge because of the scope and geographic focus of this project (e.g. state, national, or community level focus)
Are there any unique innovations and/or challenges that emerge because this project is an enhancement of an existing, and widely used, distributed research network/Enhanced Registry?

CER indicates Comparative Effectiveness Research; IRB, Institutional Review Board; PRO, patient-reported outcomes; IT, information technology; QI, Enhanced Registries for Quality Improvement.

setting, effective project coordination and management is critically important.

Project Management

The project managers' role is viewed as critical to ensuring that partnering sites remain informed and involved with one another. Clear communication and transparency regarding activities and practices at partnering sites is a major responsibility. Educating the participating sites about the infrastructure and approaches to achieve network security and privacy is important to building a "culture of trust around the technology," 1 investigator commented. Another commented, "this project is unique in that there was actually funding to develop infrastructure and governance/relationships. Before [this grant], projects had difficulty carving out portions of [grants] to do this development…if infrastructure and governance are not called out, they won't be addressed…what is durable is infrastructure, including people and relationships. We ignore these at our peril." The relatively high level of staff support for project management is

© 2012 Lippincott Williams & Wilkins

viewed as critical to success, but support for this role is unusual in the context of most research grants.

The effort to identify and disseminate best practices is another challenge for the projects. For example, project sites participating in the HMO Research Network (HMORN) can use the network's IRB and data use agreements templates and protocol guidelines for multisite studies.[4] Where these resources are not available some projects have collaboratively developed guidance to outline expected conduct for organizations and investigators. Transparency and clear guidelines are important because data partners are cautious when sharing patient-level data. Although all of the sites agreed that centralized data warehouses containing identifiable data from various sites might facilitate research, the approach would not be acceptable to data partners who want to manage the security and privacy of their own data and limit access to proprietary data. As a result, distributed and federated data networks have been developed to test the extent to which CER and QI may be conducted within the network while preserving "raw" patient-level data behind each institution's own firewall.

IRB and Other Regulatory Issues

The high time and cost burden of navigating privacy and data sharing across multiple sites and institutions[5] was discussed at length during nearly all of the site visits. Investigators at 3 sites specifically mentioned their frustration with the degree of variation in the way IRBs interpreted the regulations and conducted their review, especially with respect to the patient consent process. Project staff reported a wide range of timelines that are required to seek IRB approval, from 3 weeks–4 months. One ameliorating factor that improved the timeliness of IRB approval was the degree of involvement by project managers. These managers' efforts to coordinate multisite IRB approval through central or federated IRBs and develop new approaches to facilitate multisite data use agreements hold promising lessons for future research.

The Projects are Building Clinical Informatics Tools, Platforms, and Models to Enable Research With ECD, and it is Important to Understand the Strengths and Limitations of Each for Particular CER Questions

CER using ECD has the potential to benefit greatly from informatics tools, platforms, and models that integrate data streams across sites, data sources, and data types.[6] These strategies include a range of approaches to capture, aggregate, and integrate disparate data sources held by different institutions. Tools have also been developed to improve data access and statistical analyses. Several investigators emphasized that while no 1 informatics approach is likely to work in all settings, the lessons from each of the teams' implementation efforts will teach the community a great deal about the strengths and limitations of various informatics approaches to particular CER questions. In addition, several investigators identified the need to consider the ways that technology for CER may evolve over the course of the projects. Further information on the current literature related to the use of informatics for CER is available in an annotated bibliography produced by the EDM Forum.[7]

Range of Informatics Approaches

The 7 major informatics tools, platforms, and models that are being employed across the projects (i2b2,[8] PopMedNet,[9] TRIAD,[10] Amalga UIS,[11] RedX, DataLink, and the CER Hub[12]) support a range of end uses, including clinical decision support, operations (eg, QI), and research. Some aim to harmonize data, improve data access and exchange (by a platform, hub, and/or data marts), and improve security. Others implement middleware such as natural language processing tools to extract relevant information from free clinical text and most are integrating analytic tools for research.

Standardization and Harmonization

Because of the differences in structure and meaning of some elements in electronic health records (EHRs), efforts to harmonize data are central to the projects. To use a simple example, some EHRs use the term "sex" whereas others use "gender." However, to conduct CER, these characteristics must be harmonized. Such discrepancies in nomenclature, and potentially, in content and interpretation, are common and are important to discover and address within and across the network.

One approach to harmonizing data across multiple systems is to utilize common data models (CDM) to standardize data from partnering sites. Several projects are utilizing the HMORN Virtual Data Warehouse (VDW), which allows researchers to submit queries to the VDW and receive results from different sources. The Observational Medical Outcomes Partnership[13] CDM has also been considered by some of the projects. One team enumerated their desired set of CDM features as: availability in the public domain; past or recent field testing with similar use cases; allowing for rapid addition of new data elements; and, having an active user community.

Ontology[14] mapping systems that harmonize terminologies and definitions across multiple systems can minimize the need to standardize data across a federated network. Two projects are using ontology mapping to build a domain-specific VDW repository or patient registry to enable querying across federated networks. Given the differences in CDM and ontology-building approaches, it will be important learn from the teams' experience to understand the strengths and limitations of each for CER applications.

Improving Data Access and Statistical Analysis

Additional lessons will be learned about the range of analytic tools that several of the projects are developing to enhance research use. Particularly if working with limited datasets or deidentified data, network-based visualization and statistical tools allow researchers to gain a picture of the data that are available for research purposes and help to shape new questions. Although nearly all of the projects enable SAS code exports for further statistical analysis, 5

 © 2012 Lippincott Williams & Wilkins

projects are using or developing analytic tools that generate descriptive statistics, aggregated data tables, and visual displays of data (eg, graphs and distribution charts). Cohort identification and recruitment functions are also perceived as important to integrate into CER efforts—1 group described their efforts to build a "layered" cohorts based on diagnosis criteria from data in multiple settings (eg, outpatient office visits and pharmacy data). Two projects are implementing these functions directly in the EHR to prompt recruitment into studies or trials at the point-of-care based on desired characteristics such as diagnosis, sex, or age.

One group is working closely with their EHR vendor to get "data-in-once."[15] The data-in-once approach means that the infrastructure design should enable the ability to enter data once during front-line clinical care. This information should be sufficiently high quality to achieve a variety of uses for clinical, research, and operational needs. This approach can be valuable to integrate high-quality data collection into the workflow and into the EHR that may be used to conduct CER. At least 1 investigator felt this type of use for the EHR is a "practical choice for both the vendor and the projects involved." However, there is no agreement about the extent to which EHRs will effectively support both QI and CER, and this is another area where lessons learned will emerge over the course of the projects.

Considering the Evolving Landscape

The ever-shifting landscape of new technologies presents substantial opportunities as new devices, tools, and applications emerge. Even since the inception of the projects in 2010, the use of tablet computers and smart phones have become less expensive and more accepted for data collection. One investigator cited the availability of new technology as an important catalyst for research innovation that spurred the team to test the use of tablets for electronic data collection rather than standard paper-based instruments. Investigators anticipate that technological innovations in the coming years will benefit the projects, but also acknowledge that adapting to new technology can require substantial time and resources.

The Sites See a Need for Rigorous Methods to Assess Data Validity, Quality, and Context for Multisite Studies Given the Absence of a Gold Standard for Evaluating ECD

Bringing together disparate data sources holds tremendous potential for assessing patient experiences more fully, but also raises concerns about validity and reliability as there is no accepted gold standard or single validated source of information, such as the paper medical record,[16] which can be used to assess these data. Project staff emphasized the need for testing approaches to assure data quality[17]—which in this context focuses on accuracy, precision, and validity. At a number of sites, teams use or are developing quality assurance (QA) measures to account for human and technical errors.

At individual sites, QA is an important and ongoing effort. One project investigator commented, "if you have datasets that haven't been through a QA process you have significant challenges." In some projects these validation processes help the systems—particularly with EHR data—perform better. Some of the teams' validation approaches include efforts to model the likelihood of missing data for questions not commonly answered. Another validation approach is to integrate testing for precision to measure accuracy and level of recall in order to assess completeness of QA efforts that are wrapped into the EHR. Finally, others assess data quality at sites using informatics tools, platforms, and models, such as evaluating interrater reliability by comparing manual data abstraction with natural language processing.

For multisite studies it is also important to understand the issues that may arise when data from multiple institutions are brought together, making it possible to distinguish actual versus artifactual variability due to system level or data-quality issues. The need to test multiple dimensions of data quality at the local level is required for multisite QA to understand the level of real variability that exists across a research network; a critical element of "distinguishing variations in care from data-quality problems."[18] Other issues with data variability across sites emerge once data has been aggregated (eg, determining whether a patient whose data is unavailable after a specific date is deceased or has left the health care system).

New Opportunities Exist to Achieve Meaningful Patient and Consumer Engagement, and to Work Collaboratively in Multidisciplinary Teams

Achieving the promise of a learning health care system[19]—one that can integrate, use, and produce evidence in near-real time to improve individual and population health—requires patient and stakeholder engagement throughout the research process, and translation and dissemination strategies that may not be familiar to some researchers. The sites undertake a range of approaches to seek perspectives from relevant individuals and groups. Several projects have patient or consumer advisors on their steering committees or working groups. Other projects incorporate patient input in their technological approaches to building tools and applications, and some engage in partnerships with community centers that provide health education and opportunities for the public to directly interact with research teams. Many of the investigators expect that opportunities to bring stakeholder perspectives into the projects will expand in the future when more of the infrastructure is implemented.

At the same time, several investigators reported that they find it challenging to identify and engage the appropriate set of stakeholders—particularly patients and consumers—in CER. An expressed challenge is finding individuals who are both interested in participating and able to provide the time and meaningful input on the research. One group commented on the need to balance having stakeholders involved at the beginning with the need to establish the infrastructure first to ensure that they can demonstrate its value. Nonetheless, all projects acknowledged the importance of achieving balance and expressed interest in realizing an appropriate level of engagement.

© 2012 Lippincott Williams & Wilkins

During the site visits, many of the investigators reflected on the importance of partnership and collaboration and commented on the opportunities that emerge because of the multidisciplinary composition of the project teams, including data programmers, information technology staff, research investigators, QI professionals, clinicians, epidemiologists, biostatisticians, and other professionals. A few of the teams noted that while it can be tempting to think that information technology should be handled solely by the informaticists, with researchers focusing solely on the analytic models, the teams' experience thus far suggests it would be unwise to silo activities to such a degree. In 1 project, data programmers realized the need for multidisciplinary collaboration after experiencing trouble when they began designing informatics modules to enable the research and learned that "looking to the science provided guidance to framing the tool." In other words, building relationships across the institutions, networks, and projects may help to foster a collaborative approach that reaches across disciplinary and methodological silos to develop innovative yet practical solutions for CER.

Perhaps as a result, several teams suggested that in this "building" phase it is not necessary or possible to make clear distinctions between "infrastructure" and "science." Many of the investigators commented on the value of close working relationships between researchers, technologists, and data partners who can build systems that are capable of generating meaningful answers to priority CER questions. At 4 sites, staff emphasized that including a variety of perspectives and expertise on the team helps to bridge disciplinary gaps and spur innovation. A transdisciplinary, team-based approach that enables collaboration across areas of expertise is necessary, but requires both a willingness to bridge disciplinary silos and support for time and effort to engage.

DISCUSSION

Over the long term the project teams are aware that they are required to demonstrate the value of substantial American Recovery and Reinvestment Act investments in CER that must evolve and scale to serve a diverse set of potential users and address a range of CER or PCOR questions. To achieve this end, a high level of collaboration and support to foster partnership is necessary. As the examples highlighted in this paper demonstrate, the research project teams are exploring new approaches to resolve limitations associated with traditional research study designs and data availability.

However, the projects face tremendous time pressure to both build infrastructure and conduct CER. Building the informatics infrastructure and achieving the research objectives for each project while building partnerships and engagement strategies to ensure sustainability is an undeniable challenge in a 3 year timeframe. This pressure is more acute as the current level of federal funding provided for the projects is not likely to be repeated anytime soon. Although these realities may be daunting, EDM Forum investigators believe that the resources they are developing can be useful to a variety of audiences (eg, researchers, community members, academic institutions) and are optimistic that these new tools, data, and research products will support the case for ongoing investment.

To assist the goal of partnership and collaborative science and to build awareness of the research efforts in the PROSPECT, DRN, and Enhanced Registry programs, the EDM Forum is engaged in a number of activities. Regular webinars and annual in-person meetings support information exchange on emerging best practices for QI and research methods, as well as informatics, governance, patient and consumer engagement approaches, and clinical decision support. Strategies to ensure sustainability are also discussed. Many project members are working with the EDM Forum on commissioned papers—several of which present frameworks that will help the research community assess the strengths and limitations or particular approaches for specific CER questions. Joint outreach efforts, such as public webinars on consumer and patient engagement, or workshops on governance for multisite CER at the 2011 Public Responsibility in Medicine and Research annual conference offer an opportunity to bring the projects' lessons to key communities and decision-makers.

In the upcoming year the EDM Forum will engage the group in discussing collaborative projects and resources to improve the quality, transparency, and reproducibility of CER and PCOR. These discussions will focus on challenges and opportunities identified in this first set of site visits. Highlighting lessons learned from the ambitious and diverse projects participating in the EDM Forum is an exciting challenge on its own, the products of which will ideally enhance coordination and collaboration and ultimately generate useful and timely evidence to improve health care and population health.

ACKNOWLEDGMENTS

The authors acknowledge the helpful comments of Gurvaneet Randhawa (AHRQ), Ned Calonge (The Colorado Trust) and Michael Stoto (Georgetown University), the support of the Agency for Healthcare Research and Quality, and the editorial assistance of Beth Johnson (at AcademyHealth). Special thanks to the investigators and project staff that participated in the site visits for their assistance and helpful feedback.

© 2012 Lippincott Williams & Wilkins

APPENDIX

TABLE A1. Terms and Definitions

Term	Definition
AHRQ	Agency for Healthcare Research and Quality
CDM	Common Data Model
CDS	Clinical Decision Support
CER	Comparative Effectiveness Research
DRN	Distributed Research Network
DUA	Data Use Agreement
EDM Forum	Electronic Data Methods Forum
HMORN	Health Maintenance Organization Research Network
IRB	Institutional Review Board
IT	Information Technology
OMOP	Observational Medical Outcomes Partnership
PCOR	Patient-Centered Outcomes Research
PROSPECT	Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies studies
QA	Quality Assurance
QI	Quality Improvement
VDW	Virtual Data Warehouse
Analytic Methods	Approaches to analyzing prospective data from electronic clinical data systems for CER
Clinical Informatics	The use of computer platforms and applications to collect, process, and represent medical data used for CER
Common Data Model	An aggregated or centralized data model copies data from original sources and brings and standardizes these data in a centralized place. The copied data can then be queried and analyzed.[20]
Comparative Effectiveness Research	CER is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat, and monitor health conditions in "real-world" settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances (Patient protection and Affordable Care Act, Part D)
Data Governance	Refers to the policies and procedures that oversee data stewardship and management of data linkage, aggregation, storage, acquisition, and use of data[21]
Distributed Research Network (DRN)	A DRN is an approach in which data holders maintain control over their protected data and its uses.[22] A DRN features a central portal that performs network functions, such as operations (eg, workflow, policy rules, auditing, query formation and distribution) and security (eg, authentication, authorization) and distributed data marts that remain under the control of the data holders[23]
Federated Research Network	A federated network links geographically and organizationally separate databases so that a single database query can return results from multiple databases while maintaining the privacy and confidentiality of patient data.[24]
Learning Healthcare System	A health care system in which knowledge generation for research, science, quality assessment, outcomes, and safety standards are aligned for improvement and innovation.[19,25]
Ontology	Collects and classifies data based on relationships and content properties. Projects with an ontology build a data dictionary around a "preontology" that is used to create a queriable interchange file that links to external data in the network. New data elements are then added into the ontology and compared with existing terms, validated through review, and assigned to fields in the data dictionary.
Patient Registry	A patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes. A registry database is a file (or files) derived from the registry[26]
Patient-Centered Outcomes Research	Patient-Centered Outcomes Research helps people make informed health care decisions and allows their voice to be heard in assessing the value of health care options." It answers four patient-focused questions: "Given my personal characteristics, conditions and preferences, what should I expect will happen to me?" "What are my options and what are the benefits and harms of those options?" "What can I do to improve the outcomes that are most important to me?" "How can the health care system improve my chances of achieving the outcomes I prefer?"[27]
Rapid-cycle innovation	Rapid-cycle innovation is a method by which improvement tools and techniques can be designed and tested, with the goal of eliminating process steps that do not hold value[2]
Transdisciplinary	Transdisciplinarity connotes a research strategy that crosses many disciplinary boundaries to create a holistic approach. It applies to research efforts focused on problems that cross the boundaries of more than two disciplines, such as research on effective information systems for biomedical research, and can refer to concepts or methods that were originally developed by 1 discipline, but are now used by several others[1]
Virtual Data Warehouse (VDW)	VDW is not centralized data warehouse—it is "virtual," consisting of parallel databases setup identically at each of site that can be easily merged across sites. These databases have been constructed by extracting data from the local electronic data systems and reconfiguring them to use standard variable names and coded values[28]

© 2012 Lippincott Williams & Wilkins

REFERENCES

1. Nicolescu B. *Transdisciplinarity—Theory and Practice*. Cresskil, NJ: Hampton Press; 2008.
2. AHRQ Health Care Innovations Exchange. Adoption of Rapid Cycle Improvement Process From Toyota Increases Efficiency and Productivity at Community Health Clinics. April 14, 2008. Available at: http://www.innovations.ahrq.gov/content.aspx?id=1807. Accessed March 8, 2012.
3. Patton MG. *Qualitative Research and Evaluation Methods*. Thousand Oaks: Sage Publications Inc.; 2002:431–440.
4. HMO Research Network. HMO Research Network (HMORN). HMORN web site. 2012. Available at: http://www.hmoresearchnetwork.org/. Accessed March 8, 2012.
5. Institute of Medicine (IOM). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Washington, DC: The National Academies Press; 1991:199–244.
6. 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. *Med Care*. 2012; 50(suppl 1):S49–S59.
7. Hamilton Lopez M, Nagda SR, Holve E, et al. Comparative Effectiveness Research (CER) and Clinical Informatics: An Annotated Bibliography. AcademyHealth. 2011(EDM Forum Report). Available at: http://www.edm-forum.org/publicgrant/Publications/AnnotatedBibliography. Accessed March 8, 2012.
8. 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.
9. Harvard Pilgrim Healthcare Institute. PopMedNet. 2010. Available at: http://www.popmednet.org/Home.aspx. Accessed March 8, 2012.
10. Hastings S, Oster S, Langella S, et al. Adoption and Adaptation of caGrid for CTSA. *Summit on Translat Bioinforma*. 2009;2009:44–48.
11. Microsoft. Microsoft Amalga. 2011. Available at: http://www.microsoft.com/en-us/microsofthealth/products/microsoft-amalga.aspx. Accessed March 8, 2012.
12. Kaiser Permanente Center for Health Research. About CER Hub [CER Hub website]. 2011. Available at: https://www.cerhub.org/index.html.
13. Observational Medical Outcomes Partnership [OMOP website]. 2011. Available at: http://omop.fnih.org/. Accessed March 8, 2012.
14. 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.
15. James B. Information system concepts for quality measurement. *Med Care*. 2003;41(suppl 1):171–179.
16. 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.
17. Wang R, Kon H, Madnick S. Data Quality Requirements Analysis and Modelling. Ninth International Conference of Data Engineering, 1993.
18. Kahn MG, Raebel MA, Glanz JM, et al. A pragmatic framework for single-site and multi-site data quality assessment in electronic health recordbased clinical research. *Med Care*. 2012;50(suppl 1):S21–S29.
19. Olsen LA, Aisner D, McGinnis JM, eds. *Roundtable on Evidence-Based Medicine*. National Academies Press: Washington, DC. 2007:37–80.
20. Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. *Health Aff*. 2009;28:454–466.
21. Rosenbaum S. Data governance and stewardship: designing data stewardship entities and advancing data access. *Health Serv Res*. 2010;45:1442–1455.
22. 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(6 Suppl):S45–551.
23. D'Avolio LW, Farwell WR, Fiore LD.. Comparative effectiveness research and medical informatics. *Am J Med*. 2010;123(12 Suppl 1): e32–37.
24. 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.
25. Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. *J Clin Oncol*. 2010;28:4268–4274.
26. Gliklich RE, Dreyer NA. *Registries for Evaluating Patient Outcomes: a User's Guide*. Rockville, MD: U.S. Dept. of Health and Human Services, Public Health Service, Agency for Healthcare Research and Quality; 2007.
27. Patient-Centered Outcomes Research Institute. Working Definition of 'Patient-Centered Outcomes Research' [PCORI definition web page]. 2011. Available at: http://www.pcori.org/assets/Summary-of-PCOR-Definition-Input-Opportunity-.pdf. Accessed January 12, 2012.
28. HRN. HMORN. Virtual Data Warehouse (VDW). *Collaboration Toolkit: A guide to multicenter research in the HMO Research Network*. 2011;16–20.

 © 2012 Lippincott Williams & Wilkins