The Electronic Data Methods (EDM) Forum for Comparative Effectiveness Research (CER)

Erin Holve, PhD, MPH, MPP,†‡ Courtney Segal, BA,*†‡ Marianne Hamilton Lopez, MPA,*†‡ Alison Rein, MS,*†‡ and Beth H. Johnson, MPH*†‡*

Background: AcademyHealth convened the Electronic Data Methods (EDM) Forum to collect, synthesize, and share lessons from eleven projects that are building infrastructure and using electronic clinical data for comparative effectiveness research (CER) and patient-centered outcomes research (PCOR). This paper provides a brief review of participating projects and provides a framework of common challenges.

Methods: EDM Forum staff conducted a text review of relevant grant programs' funding opportunity announcements; projects' research plans; and available information on projects' websites. Additional information was obtained from presentations provided by each project; phone calls with project principal investigators, affiliated partners, and staff from the Agency for Healthcare Research and Quality (AHRQ); and six site visits.

Results: Projects participating in the EDM Forum are building infrastructure and developing innovative strategies to address a set of methodological, and data and informatics challenges, here identified in a common framework. The eleven networks represent more than 20 states and include a range of partnership models. Projects vary substantially in size, from 11,000 to more than 7.5 million individuals. Nearly all of the AHRQ priority populations and conditions are addressed.

Discussion: In partnership with the projects, the EDM Forum is focused on identifying and sharing lessons learned to advance the national dialogue on the use of electronic clinical data to conduct CER and PCOR. These efforts have the shared goal of addressing challenges in traditional research studies and data sources, and aim to build infrastructure and generate evidence to support a learning health care system that can improve patient outcomes.

Key Words: comparative effectiveness research, infrastructure, electronic clinical data, analytic methods, clinical informatics, data governance, learning health care system

(*Med Care* 2012;50: S7–S10)

From the *Electronic Data Methods Forum, AcademyHealth; †Research and Education in Health Services Research, AcademyHealth; and ‡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.

ISSN: 0025-7079/12/5007-00S7

REPORT

Improving the relevance, rigor, and timeliness of evidence on the comparative effectiveness of health care treatments and strategies is a critical goal for comparative effectiveness research (CER).[1] As a part of the $1.1 billion in support for CER through the American Recovery and Reinvestment Act, approximately $100 million was provided to build a more robust and flexible infrastructure for conducting CER using prospective electronic clinical data.[2] With support from the Agency for Healthcare Research and Quality (AHRQ), AcademyHealth has convened the Electronic Data Methods (EDM) Forum to identify, synthesize, and share lessons learned from 11 projects that are building infrastructure and using electronic clinical data for CER and patient-centered outcomes research (PCOR). In its activities, the EDM Forum considers 4 domains: analytic methods, clinical informatics, data governance, and CER and the learning health care system.

This report provides a brief review of the research networks participating in the EDM Forum and is based on an environmental scan conducted by the EDM Forum. EDM Forum staff conducted a text review of the AHRQ funding opportunity announcements for the 3 grant programs,[3–5] research plans of the 11 research projects, and information available on the projects' websites (if applicable), and attended overview presentations of each project (facilitated by the EDM Forum). These efforts contributed to a deeper understanding of the projects participating in the EDM Forum. In some instances, EDM Forum staff participated in phone calls with project principal investigators, affiliated partners, and AHRQ staff (including the project officer), and synthesized information was confirmed and/or refined during a set of site visits conducted at 6 of the projects in the spring of 2011.

As described in Dr Gurvaneet Randhawa's a report in this special issue, the projects are supported through 3 grant programs, The Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies (PROSPECT), Scalable Distributed Research Network (DRN) for CER, and the Enhanced Registry for Quality Improvement (QI) and CER projects. An esteemed, multidisciplinary group of researchers are engaged in these projects. An organizational chart and full list of project titles and investigators is available at http://www.edm-forum.org.[6]

The goal of these research programs is to build a flexible infrastructure that can be leveraged to address some of the common problems in traditional research studies (both

TABLE 1. Framework of the Challenges of Traditional Research Studies and Electronic Clinical Data for Specific Attributes Necessary to Conduct CER

	Methodological Challenges		Data and Informatics Challenges	
	Randomized Controlled Trials	**Observational Studies**	**Claims-based Data**	**Clinical Data Collected at the Point-of-Care**
Meets specific requirements of CER	Generally focused on efficacy May not be sufficiently responsive to changes in practice Rare outcomes are generally not captured (because of sample size issue) Safety outcomes are not routinely or completely captured Strict protocols may prevent reporting of adverse events from drug-drug interactions that are seen in real-world practice	Limited ability to identify causal relationships as assignment is not randomized	Created for billing; often Limited clinical information Lacking detail on diagnostic tests and results Static databases do not permit rapid updates or modifications	To the extent data is deidentified, may not have ability to: access sufficiently granular data access specific dates of service for distributed data, ability to pool data may be limited by DUAs, possibly limiting ability to answer some CER questions Current systems are often not optimized to identify potential enrollees for pragmatic trials
Data quality	Differential loss to follow-up and other information biases that could bias study results	Lack of randomization results in selection bias (eg, treatment by indication or susceptibility bias) Missing data is prevalent (patient or item level)	Often hard to unbundle claims Coding may not be implemented consistently Potential for upcoding Potential for missing data related to denied claims	Lack of "research grade" data in EHRs Lack of data on confounders needed to construct subgroups Unstructured data capture/entry Semantic interoperability (no standard vocabulary/ontology) Limited compliance with data transfer standards Virtually nonexistent standardization of data across vendors
Feasibility/ practicality	Generally not feasible for long-term outcomes Number of comparator groups that (for feasibility/scalability reasons) can be simultaneously considered are limited because of power and multiple comparisons concerns Generally limited to 1 outcome of interest Often expensive; one-off studies	Large prospective cohort studies may be just as expensive as some small RCTs or pragmatic trials	May be expensive to access granular data for CER Limited information (see CER requirements above)	Potential lack of flexibility to add new items or concepts to EHR Ability to engage participation from data partners
Generalizability/ representative-ness	Highly selected patient populations Controlled clinical settings and teams Often only short-term or surrogate efficacy outcomes are measured Limited sample size prevents ascertaining outcomes in all relevant subgroups	Variation in cohort definitions and prognostic factors are prevalent that can impact reproducibility	Difficult to harmonize across multiple sources and settings Limited ability to link streams of data	Limited ability to link streams of data (eg, PRO to EHR)
Timeliness	Not suitable for many decision makers (clinical and policy) because of: rapidly evolving technology need to assess a large number of inputs/variables Depending on condition and intervention: enrollment can take a long time detection of key interim and longer-term outcomes/markers can take a long time Often 4–7 y lag between approval, conduct, and publication of results	Combination of data lag and time to publication can significantly impact timeliness (often 3–5 y lag)	Data obtained after considerable time (often 2–3 y lag) if researchers do not have access behind organizations' firewalls	Potential for faster turnaround time, but it is unclear how rapidly EHR data can and should be analyzed for clinical decision-making (eg, what level of processing/validation is required?) For specific purposes (eg, quality improvement) timeliness may be more important; effective data use agreements are critical to ensure data can flow between sites

CER indicates comparative effectiveness research; DUA, data use agreement; EHR, electronic health record; PRO, patient-reported outcomes; RCT, randomized control trial.

randomized controlled trials and observational studies) and data sources (administrative claims data and electronic health records).[3–5] Table 1 provides a framework of common challenges for CER that have been enumerated by grantees and stakeholders. The table addresses the limitations of traditional randomized control trials and observational studies, as well as the use of claims-based data and clinical data collected at the point-of-care (eg, electronic health records) for CER.

To harness the potential of electronic clinical data or CER, each of the projects is blending data streams from a variety of institutions and data types including administrative, clinical, and patient-reported data. These data will develop a fuller picture of patient experience that can be used to

inform research and quality improvement efforts. Collaboration among Federal agencies; organizations at the Federal, state and local levels; and between the public and private sectors is also critical.[7] As a result, all of the projects are investing in multi-institution partnerships and developing transdisciplinary teams to ensure sustainable infrastructure and processes are in place to allow "data to flow" for research, QI, and clinical decision support.

The projects represent a range of partnership models and geographic scope, and vary substantially in size. Some are national networks, whereas others are more locally focused. Two of the 11 projects are state-specific (Washington and Indiana) and 1 project focuses locally on a specific community in New York City. Overall, 20 states are represented. The networks also range in size from 11,000 to 7.5 million individuals. The estimates of network size that are included in this report reflect the number of patients with specific priority populations and conditions who would be eligible to participate in the CER studies conducted by the projects. In a few cases, the size of the full network is much larger, specifically with respect to the projects that are affiliated with the HMO Research Network (HMORN). In total, the number of patients included in the HMORN is estimated to include more than 18 million patients. Figure 1 represents the geographic scope of all the 11 projects participating in the EDM Forum.

In addition to building infrastructure for large, networked studies, each of the projects includes a set of specific research studies to demonstrate proof of concept for using electronic clinical data for CER and PCOR. Thirty-eight such CER studies are planned or underway on a range of populations and topics. As a whole, the studies address nearly all of the AHRQ priority populations and conditions, with several projects focused on low-income populations, racial and ethnic minorities, and topical areas such as hypertension and asthma. Table 2 provides a summary of the number of studies addressing specific priority populations and conditions. The studies are also conducting CER for a range of treatments and interventions, diagnosis and adherence measures, care management strategies, care delivery characteristics, and quality of life measures. Patient-reported outcomes are being collected and analyzed in most of the projects.

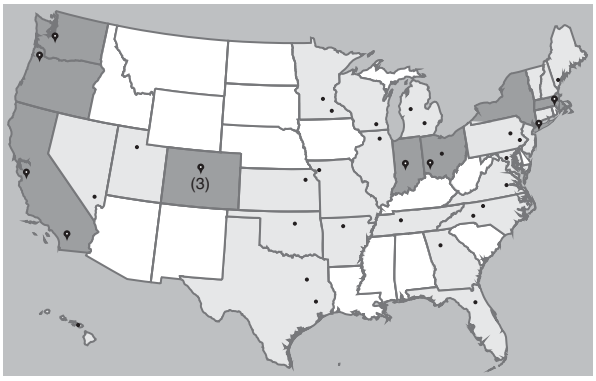

FIGURE 1. Geographic Scope of the PROSPECT, DRN, and Enhanced Registry Studies, 2012. This map represents the EDM Forum network of projects building the electronic clinical data infrastructure to improve patient outcomes. Coordinating sites of each of the 11 research projects are noted with large markers. Three coordinating sites are located in Denver, CO. Participating sites in the research networks (ie, health care settings, research sites, and data sharing partners) are noted with small points. *Shading Legend*: Dark shading indicates $\geq$ 20% of the state's population is included in at least one of the networks; light shading indicates between 2% and up to 20% of the state's population is included in at least one of the networks; no shading indicates that there is a not a significant number of patients from the research networks represented in the state.

TABLE 2. Number of Comparative Effectiveness Research Studies in the PROSPECT, DRN, and Enhanced Registry Projects That Address Specific AHRQ Priority Populations or Conditions

	# CER Studies
AHRQ priority populations	
Racial and ethnic minorities	5
Low-income	6
Inner-city	3
Rural	1
Children	3
Elderly	3
People with chronic conditions	3
Women	4
AHRQ priority conditions	
Attention Deficit Hyperactivity Disorder	1
Asthma	4
Alzheimer	1
Chronic obstructive sleep apnea	1
Diabetes (I and II, and gestational diabetes)	2
Hypercholesterolemia	1
Hypertension	5
Obesity	3
Pediatric inflammatory bowel disease	1
Peripheral vascular disease	1
Substance abuse (smoking cessation)	1

AHRQ indicates Agency for Healthcare Research and Quality; CER, comparative effectiveness research; DRN, Distributed Research Network; PROSPECT, Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies.

Building infrastructure to conduct CER and PCOR using electronic clinical data has been a growing enterprise for more than a decade. The following list of projects encompasses major efforts to build research infrastructure with electronic health data: Clinical and Translational Science Awards (CTSA),[8–11] including workgroups, systems, and tools developed through the CTSA; Health Maintenance Organization Research Network[12]; Distributed Ambulatory Research in Therapeutics Network[13]; Food and Drug Administration (FDA) Sentinel System[14]; FDA Partnership in Applied Comparative Effectiveness Science[15]; FDA Janus Clinical Trial Data Repository[16]; Observational Medical Outcomes Partnership[17]; Office of the National Coordinator Beacon Community Program[18]; Veterans Affairs Informatics and Computing Infrastructure[19]; Strategic Health IT Advanced Research Projects Program[20]; QueryHealth[21]; and The Shared Health Research Information Network.[22]

The EDM Forum's efforts with the PROSPECT, DRN, Enhanced Registry projects share a common goal with many

of these initiatives—to generate evidence and build a learning health care system that can improve patient outcomes. A critical aspect of this charge is identifying lessons learned from relevant initiatives to advance the national dialog on the use of electronic clinical data for the conduct of CER, QI, and clinical decision support. The EDM Forum supports these efforts by fostering exchange of information on promising practices, and identifying opportunities for collaboration across the research project teams, related activities, and key stakeholders across a variety of organizations and agencies.

REFERENCES

1. US Department of Health and Human Services. Text of the Recovery Act Related to Comparative Effectiveness Funding. Excerpt from the American Recovery and Reinvestment Act of 2009. [HHS Recovery Act web site]. 2009. Available at: http://www.hhs.gov/recovery/programs/cer/recoveryacttext.html. Accessed August 31, 2011.
2. US Department of Health and Human Services. Comparative Effectiveness Research Funding. [HHS Recovery Act web site]. 2010. Available at: http://www.hhs.gov/recovery/programs/cer/index.html. Accessed August 31, 2011.
3. Agency for Healthcare Research and Quality. ARRA-AHRQ Recovery Act 2009 Limited Competition: PROSPECT Studies: Building New Clinical Infrastructure for Comparative Effectiveness Research (R01). [PROSPECT Funding Opportunity Announcement web site]. 2009. Available at: http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-10-005.html. Accessed November 1, 2011.
4. Agency for Healthcare Research and Quality. ARRA OS Recovery Act 2009 Limited Competition: Enhanced Registries for Quality Improvement and Comparative Effectiveness Research (R01) [Enhanced Registries Funding Opportunity Announcement web site]. 2010. Available at: http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-10-020.html. Accessed November 1, 2011.
5. Agency for Healthcare Research and Quality. ARRA OS: Recovery Act 2009 Limited Competition: Scalable Distributed Research Networks for Comparative Effectiveness Research (R01). [Scalable Distributed Research Networks Funding Opportunity Announcement web site]. 2010. Available at: http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-10-015.html. Accessed November 1, 2011.
6. AcademyHealth. EDM Forum Projects & Investigators [EDM Forum web site]. 2011. Available at: http://www.edmforumresearchportal.org/publicgrant/About/projectprofiles/. Accessed November 1, 2011.
7. Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and The Congress: US Department of Health and Human Services; June 2009.
8. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–381.
9. Hastings S, Oster S, Langella S, et al. Adoption and Adaptation of caGrid for CTSA. *Summit on Translat Bioinforma*. 2009;2009:44–48.
10. 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.
11. Payne P, Ervin D, Dhaval R, et al. TRIAD: The Translational Research Informatics and Data Management Grid. *Appl Clin Inform*. 2011;2:331–344.
12. HMORN HRN. Virtual Data Warehouse (VDW). *Collaboration Toolkit: A Guide to Multicenter Research in the HMO Research Network*. 2011:16–20.
13. Pace W, West D, Valuck R, et al. *Distributed Ambulatory Research in Therapeutics Network (DARTNet): Summary Report*. Rockville, MD: Agency for Healthcare Research and Quality; 2009.
14. U.S. Food and Drug Administration. Sentinel Initiative. [Sentinel web site]. 2012. Available at: http://www.fda.gov/Safety/FDAsSentinelInitiative/. Accessed February 22, 2012.
15. Food and Drug Administration. Partnership in Applied Comparative Effectiveness Science (PACES). [PACES Solication web page]. 2010. Available at: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=118a867ca5850ebbcd3a06961355b285&_cview=0. Accessed March 2, 2012.
16. Food and Drug Administration. Janus Operational Pilot. [Janus web page]. 2009. Available at: http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm155327.htm. Accessed March 3, 2012.
17. Foundation for the National Institutes of Health. Observational Medical Outcomes Partnership. [OMOP web site]. 2011. Available at: http://omop.fnih.org. Accessed February 22, 2012.
18. Maxson ER, Jain SH, McKethan AN, et al. Beacon communities aim to use health information technology to transform the delivery of care. *Health Aff (Millwood)*. 2010;29:1671–1677.
19. U.S. Department of Veterans Affairs. VA Informatics and Computing Infrastructure (VINCI). [VINCI web site]. 2012. Available at: http://www.hsrd.research.va.gov/for_researchers/vinci/. Accessed February 22, 2012.
20. Office of the National Coordinator for Health Information Technology. Strategic Health IT Advanced Research Projects (SHARP) Program. [SHARP web site]. 2011. Available at: http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__sharp_program/. Accessed February 22, 2012.
21. S&I Framework. Query Health. [Query Health wiki]. 2012. Available at: http://wiki.siframework.org/Query+Health. Accessed February 22, 2012.
22. Weber GM, Murphy SN, McMurry AJ, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc*. 2009;16:624–630.