Developing a data infrastructure for a learning health system: the PORTAL network

Elizabeth A McGlynn,[1] Tracy A Lieu,[2] Mary L Durham,[3] Alan Bauck,[3] Reesa Laws,[3] Alan S Go,[2] Jersey Chen,[4] Heather Spencer Feigelson,[5] Douglas A Corley,[2] Deborah Rohm Young,[6] Andrew F Nelson,[7] Arthur J Davidson,[8] Leo S Morales,[9] Michael G Kahn[10,11]

For numbered affiliations see end of article.

Correspondence to
Dr Michael G Kahn, University of Colorado, Building 500, Mail Stop 563, 13001 East 17th Place, Room C1009, Aurora, CO 80045, USA; Michael.Kahn@ucdenver.edu

ABSTRACT

The Kaiser Permanente & Strategic Partners Patient Outcomes Research To Advance Learning (PORTAL) network engages four healthcare delivery systems (Kaiser Permanente, Group Health Cooperative, HealthPartners, and Denver Health) and their affiliated research centers to create a new national network infrastructure that builds on existing relationships among these institutions. PORTAL is enhancing its current capabilities by expanding the scope of the common data model, paying particular attention to incorporating patient-reported data more systematically, implementing new multi-site data governance procedures, and integrating the PCORnet PopMedNet platform across our research centers. PORTAL is partnering with clinical research and patient experts to create cohorts of patients with a common diagnosis (colorectal cancer), a rare diagnosis (adolescents and adults with severe congenital heart disease), and adults who are overweight or obese, including those with pre-diabetes or diabetes, to conduct large-scale observational comparative effectiveness research and pragmatic clinical trials across diverse clinical care settings.

BACKGROUND

Kaiser Permanente & Strategic Partners Patient Outcomes Research To Advance Learning (PORTAL) is a network that brings together four leading healthcare delivery systems: Kaiser Permanente, Group Health Cooperative, HealthPartners, and Denver Health. These systems include 11 affiliated research centers and serve 11 million members across nine states and the District of Columbia (table 1; figure 1), or approximately one of every 30 people in the USA. The four PORTAL health systems own or operate 44 hospitals, 674 clinics or medical offices, and 625 in-house pharmacies. The PORTAL partners represent a great diversity of care models and practices. The Kaiser Permanente regions, for example, contract out only a small proportion of ambulatory care, inpatient care, imaging, and highly specialized tertiary services. Group Health Cooperative, on the other hand, contracts externally for a relatively large proportion of its physicians and facilities services.

THE PORTAL DATA ENVIRONMENT

PORTAL network sites represent large and diverse integrated health systems that include inpatient and outpatient facilities; primary care and specialty provider networks; and ancillary services, pharmacies, and ambulatory procedure centers. Detailed clinical and administrative data are integrated into a comprehensive longitudinal electronic health record (EHR). Each participating PORTAL organization maintains a separate electronic record that captures in-system encounters and incorporates billing and clinical care data from services delivered by providers outside the organization. Extensive historical records and relatively long average enrollment periods allow PORTAL sites to participate in comparative effectiveness research (CER) studies that involve long periods of time for exposure or outcomes.

The PORTAL common data model

Collaboration to conduct research among these different health plans and organizations presents challenges. PORTAL healthcare systems use different EHR vendors and different clinical, administrative, and patient-access applications to support clinical care. Institutional differences in configurations, workflows, and codes create additional barriers to sharing data *directly* using existing systems. Even among partners who use EHRs from the same vendor, differences in products, capabilities, versions, and local configurations create dissimilarities in data variable names, formats, and meanings.

To address these issues, PORTAL will use a common data model (CDM) that provides definitions for how each shared data element must be structured and which codes must be assigned to data values.[1] CDMs have been used successfully in large-scale national data sharing networks.[2–4] One data model used in multiple national networks is the HMO Research Network Virtual Data Warehouse (HMORN VDW).[5][6] Over a 20-year period, the HMORN has developed detailed definitions, documentation, and implementation guides for the structure of each data table and the allowed codes used in each field.[6][7] The HMORN has developed an extensive set of data validation routines to assess data quality in VDW data extractions.[8]

The Kaiser Permanente Center for Effectiveness and Safety Research (CESR) CDM is an *expansion* of the current HMORN VDW and is the data model that will be implemented across PORTAL sites. Figure 2 illustrates the data domains defined in the current and future versions of the CESR CDM. A critical priority of the Patient Centered Outcomes Research Institute (PCORI) is to include a wide range of patient-reported outcomes (PROs) in addition to traditional clinical and administrative data. The CESR CDM contains four additional

Table 1 The PORTAL health systems

Kaiser Permanente (KP)	KP was founded in 1945 to provide high-quality, affordable healthcare services and to improve the health of its members and the communities it serves. The KP Medical Care Program is an integrated delivery system comprising more than 9 million members in seven regions in eight states and the District of Columbia. Care for members and patients is focused on their total health and guided by their personal physicians, specialists, and teams of caregivers.
Group Health Cooperative (GHC)	GHC, located in Washington State, began in 1947 as a community coalition dedicated to making quality healthcare available and affordable. Along with HealthPartners, GHC is one of the few healthcare organizations in the country governed by consumers. Its 11-member Board of Trustees—all health plan members elected by other members—work closely with management and medical staff to ensure that the organization's policies and direction put the needs of patients first. GHC seeks to promote patient-centered care and innovation by continually focusing on the needs of its 407 000 members.
Health Partners (HP)	HP was founded in 1957 in Minnesota as a cooperative. Today, HP is the largest consumer-governed, non-profit healthcare organization in the nation. In 2013, HP partnered with Park Nicollet, a non-profit, integrated healthcare system in Minnesota, to improve health, patient experiences, and affordable care, and engage members and the community. The two organizations are now officially joined under the name HealthPartners and a single, consumer-governed board of directors. This new integrated healthcare and financing organization serves more than 1.4 million medical and more than 1 million dental patients in Minnesota and western Wisconsin.
Denver Health (DH)	DH was established in 1860 and is a comprehensive, integrated organization providing level-one care, regardless of ability to pay. Twenty-five percent of all Denver residents, or approximately 130 000 individuals, receive their healthcare at DH. DH cares for one in three children in Denver. As Colorado's primary safety net institution, DH has provided billions of dollars in uncompensated care. DH is an integrated, efficient, high-quality healthcare system serving as a model for other safety net institutions across the nation.

Patient Reported Outcomes tables for storing patient-reported data such as exercise as a vital sign (EVS), the Patient Health Questionnaire (PHQ-9),[9] and the Brief Pain Inventory Survey (BPI).[10] These new tables ensure that PORTAL will be able to store and analyze PROs in alignment with a central PCORI objective.

Information exchange, both among PORTAL sites and between the PORTAL network and other PCORnet networks, requires both syntactic (structure) and semantic (meaning) harmonization.[11 12] For sharing data between multiple networks, mappings between CDMs can be constructed to provide syntactic harmonization. Semantic harmonization, however, can be difficult if two networks use different terms, coding systems, and data definitions.[13–16] While not a complete solution to full

semantic harmonization, the CESR data model uses widely adopted national and international coding systems as data elements and values (table 2, figure 3). In August 2012, the Centers for Medicare and Medicaid Services and the Office of the National Coordinator for Health Information Technology released the Final Rule for Meaningful Use Stage 2, which specifies multiple terminologies that must be incorporated into certified EHR products by October 2014.[17–19] The CESR CDM contains all of the terminologies specified in these regulations except for SNOMED Clinical Terms (SNOMED CT). As the delivery systems at PORTAL sites transition to these new coding systems, the CESR CDM will extract data elements encoded in these new terminologies. Because the current CESR CDM has the ability to record data in multiple coding standards
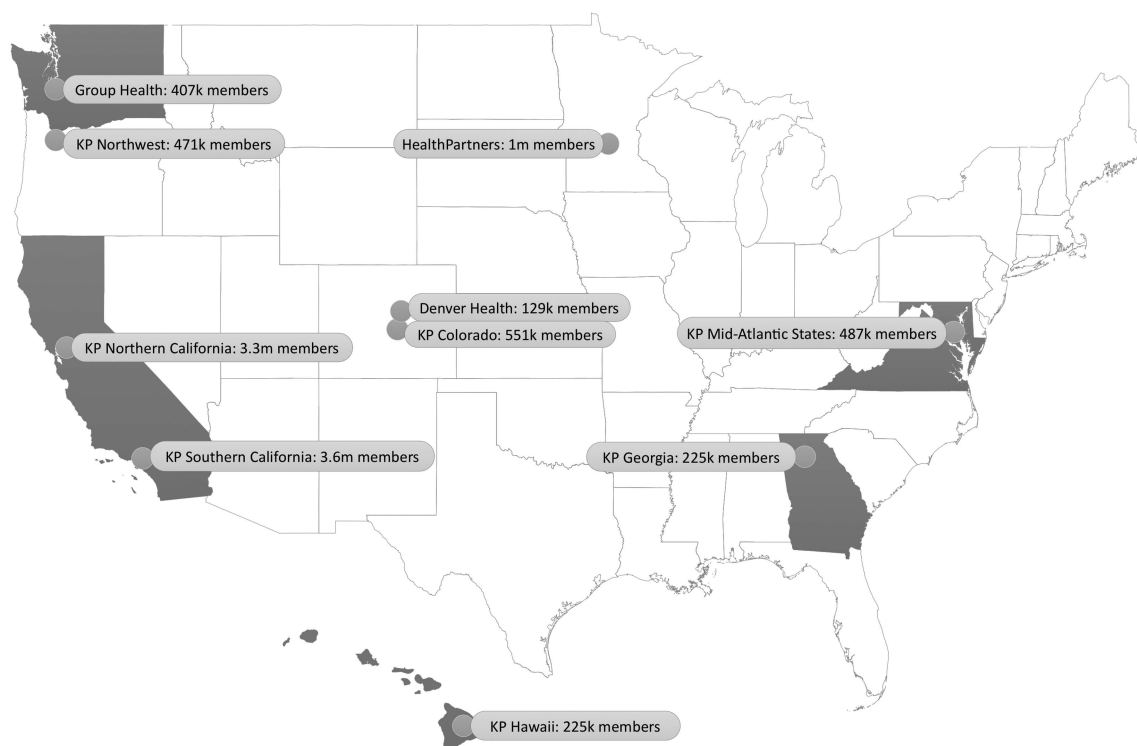
Figure 1 Geographic distribution of PORTAL clinical practices. KP, Kaiser Permanente; k, thousands(s); m, million(s).

Figure 2 The Kaiser Permanente
Center for Effectiveness and Safety
Research (CESR) common data model.
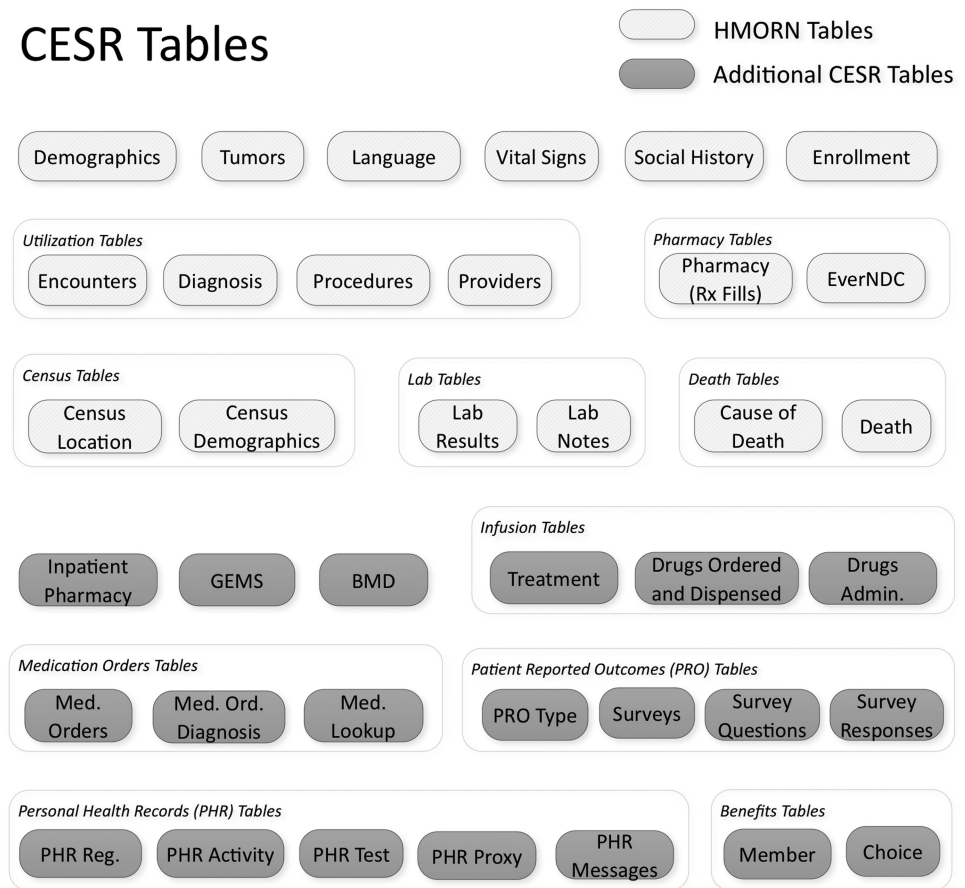
Table 2 National/international terminology standards used in the
Kaiser Permanente Center for Effectiveness and Safety Research
(CESR) common data model

Table/clinical domain	National/international coding standards in CESR data model
Demographics	Race: NIH*
Encounters	CMS DRG
Diagnosis	ICD-7-CM, ICD-8-CM, ICD-9-CM, ICD-10-CM
Procedures	ICD-9, ICD-10, ICD-11, CPT-3, CPT-4, HCPCS-3, HCPCS-4†
Tumor	ICD-O-3‡, SEER SS1997, SEER SS2000§, Facility Oncology Registry Data Standards¶
Pharmacy dispensings	NDC, Medi-Span GPI**, AHFR Pharmacologic-Therapeutic Classification System††
Medication orders	ICD-9, ICD-10 diagnoses associated with medication orders
Census	FIPS/NIST geocoding standards‡‡
Death	ICD-9, ICD-10
Laboratory	LOINC

URLs to less common coding standards are provided as follows:
*http://www.whitehouse.gov/omb/fedreg_1997standards
†http://www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/
TransactionCodeSetsStands/index.html
‡http://www.who.int/classifications/icd/adaptations/oncology/en/
§http://seer.cancer.gov/tools/ssm/
¶http://www.facs.org/cancer/coc/programstandards2012.html
**http://www.medispan.com/uniform-system-of-classification-cross-reference-file/
††http://www.ahfsdruginformation.com/license-pt-classification.aspx
‡‡http://www.nist.gov/itl/fipsinfo.cfm

(eg, diagnosis in both ICD and SNOMED CT), adding
SNOMED CT as valid values for coded data elements will not
require changes to the data model structure. This functionality
permits the co-existence of legacy data in legacy coding systems
and data captured using newer coding systems, which is critical
for conducting long-term longitudinal observational CER
studies.

Distributed data sharing platform

Distributed data queries and data exchange across PORTAL
partners will be managed using PopMedNet (http://www.
popmednet.org) technology (figure 3).[20] PopMedNet provides
the security, authentication, and auditing capabilities required to
ensure only approved data requests are submitted and returned.
PopMedNet is a data-model agnostic distributed data-sharing
platform that supports a wide range of data governance models.
PCORI has selected PopMedNet to support PCORnet's
network-of-networks data sharing infrastructure.

Ensuring data consistency and quality across network partners

The PORTAL network will build upon the experiences of other
established networks to develop new partnerships. Over the past
20 years, for example, the HMORN has developed extensive
policies, procedures, and technologies for evaluating and investi-
gating data validation, quality, and consistency. Brown and
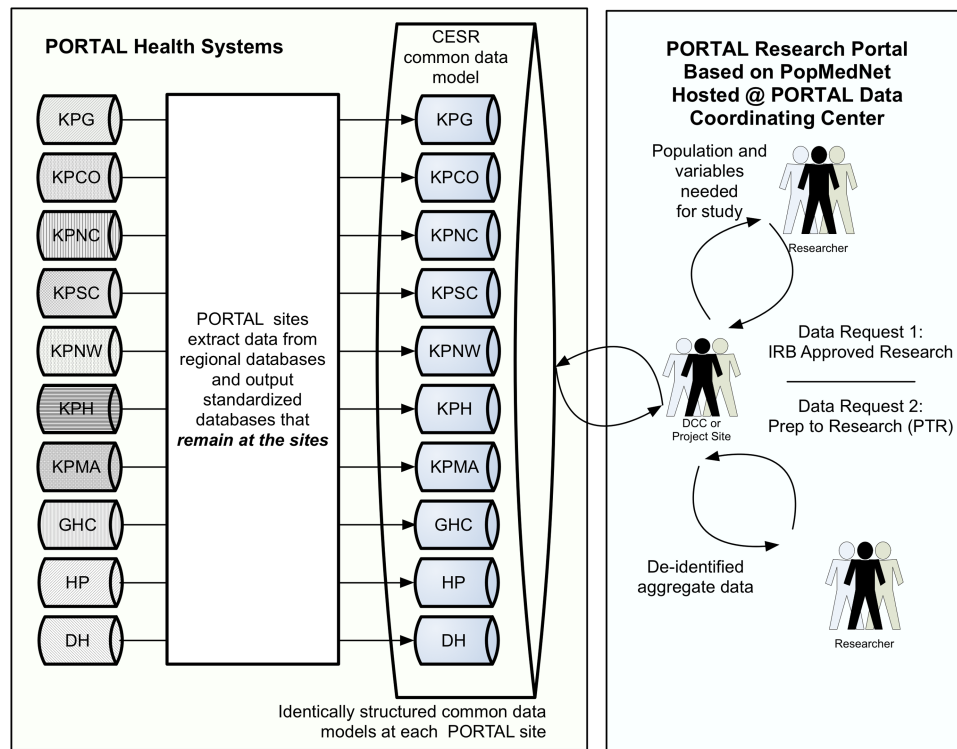
Figure 3 Data integration via the Kaiser Permanente Center for Effectiveness and Safety Research (CESR) common data model and PopMedNet distributed query platform. DH, Denver Health; GHC, Group Health Cooperative; HP, Health Partners; KP {G, CO, NC, SC, NW, H, MA}, Kaiser Permanente {Georgia, Colorado, Northern California, Southern California, Northwest, Hawaii, Mid-Altantic}.

colleagues illustrate some of the 'lessons learned' from the vast field experience within the HMORN and Mini-Sentinel networks.[21] Additionally, Kahn has published a detailed data quality assessment (DQA) model[22] and, in collaboration with the Electronic Data Methods Forum, has developed a set of recommendations for standardized DQA reporting measures.[23] Similarly, Bauck and colleagues developed a conceptual model for a consistent DQA framework that is being implemented across the HMORN/CESR sites.[8] PORTAL will draw upon these sources when developing common DQA policies and procedures and common data quality output structures so that investigators seeking to combine data from multiple networks can evaluate data quality measures from each participating site to assess their 'fitness for use' for their research question prior to incorporating data from that site.

PORTAL COHORTS
To ensure broad applicability, PCORI required each research network to develop cohorts representing a common clinical condition and a rare clinical condition. All networks were also required to develop an obesity cohort. The PORTAL network will construct three cohorts: (1) patients with a diagnosis of colorectal cancer (CRC), representing a common disease; (2) adolescents and adults with severe congenital heart disease (CHD), representing a rare disease; and (3) adults who are overweight or obese, including those who have pre-diabetes or diabetes. The characteristics of these cohorts are described briefly.

Colorectal cancer
We chose CRC because it is the third most common cancer in the USA, is the second leading cause of cancer death, and affects both men and women. Approximately 1.2 million people in the

USA currently live with CRC, which offers opportunities for studying issues of survivorship, including cancer treatment and transitions in care between primary and specialty physicians (eg, primary care, surgery, gastroenterology, and oncology). This allows researchers to examine differences in screening, treatment choices, and survivorship experiences by gender,[24] race/ethnicity, comorbid conditions, and patient preferences. There are more than 11 000 individuals with CRC across the network.

Severe congenital heart disease
Adolescents and adults with severe CHD were selected because this group faces three generalizable challenges to healthcare systems: (1) transitions in care from adolescence to adulthood; (2) monitoring of patients at increased risk for chronic conditions and associated morbidity and mortality (specifically, chronic heart failure); and (3) interfaces between primary, specialty (general cardiology), and subspecialty (CHD-specific) care. The Centers for Disease Control and Prevention recently convened a panel of experts that identified two gaps in understanding the public health implications of this condition: long-term outcomes for persons with CHD and the appropriateness of care delivery, particularly through the transition from adolescence to adulthood.[25] The PORTAL network contains approximately 330 patients with severe CHD.

Obesity
More than one-third of adults in the USA are obese.[26] The prevalence of obesity is similar for men and women, more common among persons age 60 and older, and varies by race/ethnicity, with non-Hispanic black individuals having the highest age-adjusted rates of obesity (49.5%). Obesity's relationship to diabetes is well established, with more than 10% of the US adult population currently diagnosed with diabetes and with

a prevalence greater than 25% for adults over the age of 65 years. Another 79 million adults have pre-diabetes, a condition of abnormally high blood glucose levels and a precursor to diabetes. Compared with white individuals, Mexican Americans and black individuals have a 87% and 77% higher risk of developing diabetes, respectively.[26] Each of the clinical data research networks will develop a cohort of persons who are overweight or obese that will demonstrate PORTAL's ability to work across the network of networks that comprise PCORnet. The PORTAL network has over 3 000 000 individuals who meet the criteria for obesity.

INCORPORATING PATIENT-REPORTED DATA IN ROUTINE CLINICAL PRACTICE

PORTAL members recognized the importance of capturing patient-reported data directly into the EHR many years ago. Three measures are routinely collected at all Kaiser Permanente and Group Health Cooperative sites: EVS, the BPI, and the PHQ-9, making these variables available to investigators seeking to link patient outcome measures to disease states, therapeutic interventions, and clinical outcomes.

PORTAL members have identified six critical success factors/barriers for incorporating patient-reported data into routine care delivery. First, clinicians are more likely to adopt and use measures that enhance the clinician's ability to deliver high-quality care. Clinicians often see disease-specific measures, such as the PHQ-9, as more relevant than general measures of overall functional status. Second, data collection must be hard-wired into daily workflows to ensure complete data capture. EVS measures, for example, are integrated into the routine information gathering performed by the medical assistant or nurse during the visit intake process. Third, resources must be available to ensure that the necessary functionality is implemented and is consistent with regulatory and compliance requirements. Fourth, the placement of information in the clinical record must be convenient and interpretable. All too often, PROs appear as a separate tab in the record or as a PDF that must be selected separately to view. Fifth, in some instances, patients have been reluctant to have these data incorporated into their medical record. For example, only 65% of members who take a Total Health Assessment Survey through Kaiser Permanent's web portal agree to share this information with their physician. Sixth, patients' willingness to provide this information depends on their belief that the data will be used in practice. These six principles, gleaned from many years of experience with a wide range of measures, will guide PORTAL's development of a sustainable data collection strategy within routine clinical practice.

CONCLUSION

With nearly 11 million people and more than 15 years of collaborative history among most of its partner sites, the PORTAL network offers a robust and experienced platform for comparative effectiveness and patient-centered outcomes research. This network holds promise for enhancing, storing, and analyzing patient-reported data and adopting new approaches for patient, clinician, and stakeholder engagement in all aspects of research, from the development of high-impact questions to the design of interventions and data collection approaches. These results will ultimately improve healthcare practices. As a partner in PCORnet, PORTAL can be a significant contributor to, and benefactor from, the rapidly evolving new model for interoperable large-scale national collaborative patient-centered research networks.

Author affiliations
[1]Kaiser Permanente Center for Effectiveness and Safety Research, Pasadena, California, USA
[2]Division of Research, Kaiser Permanente Northern California, Oakland, California, USA
[3]Center for Health Research, Kaiser Permanente Northwest, Portland, Oregon, USA
[4]Mid-Atlantic Permanente Research Institute, Kaiser Permanente Mid-Atlantic States, Rockville, Maryland, USA
[5]Institute for Health Research, Kaiser Permanente Colorado, Denver, Colorado, USA
[6]Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, California, USA
[7]HealthPartners Institute for Education and Research, Minneapolis, Minnesota, USA
[8]Denver Public Health, Denver Health, Denver, Colorado, USA
[9]Group Health Cooperative, Group Health Research Institute, Seattle, Washington, USA
[10]Department of Pediatrics, University of Colorado, Aurora, Colorado, USA
[11]Colorado Clinical and Translational Sciences Institute, Aurora, Colorado, USA

REFERENCES

1 Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. *Med Care* 2012;50(Suppl):S60–7.
2 Brown JS, Lane K, Moore K, *et al*. Defining and evaluating possible database models to implement the FDA Sentinel initiative. 2009. http://www.regulations.gov/contentStreamer?objectId=0900006480e6cd93&disposition=attachment&contentType=pdf (accessed 15 Feb 2014).
3 Pace WD, West DR, Valuck RJ, *et al*. Distributed Ambulatory Research in Therapeutics Network (DARTNet): Summary Report. 2009. http://effectivehealthcare.ahrq.gov/ehc/products/53/151/2009_0728DEcIDE_DARTNet.pdf (accessed 15 Feb 2014).
4 Brown JS, Holmes JH, Shah K, *et al*. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care* 2010;48:S45–51.
5 Platt R, Davis R, Finkelstein J, *et al*. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on Therapeutics. *Pharmacoepidemiol Drug Saf* 2001;10:373–7.
6 Anonymous. HMO Research Network. Top Tools & Materials. http://www.hmoresearchnetwork.org/en/ (accessed 21 Apr 2014).
7 Anonymous. HMO Research Network- VDW Data Model. http://www.hmoresearchnetwork.org/en/Tools%20&%20Materials/VDW/ (accessed 12 Feb 2014).
8 Bauck A, Bachman D, Riedlinger K, *et al*. Developing a Structure for Programmatic Quality Assurance Checks on the Virtual Data Warehouse [abstract]. Clinical Medicine & Research 2011;9:184.
9 Center for Quality Assessment and Improvement in Mental Health. The Patient Health Questionnaire (PHQ-9)—Overview. http://www.cqaimh.org/pdf/tool_phq9.pdf (accessed 15 Mar 2014).
10 Tan G, Jensen MP, Thornby JI, *et al*. Validation of the Brief Pain Inventory for chronic nonmalignant pain. *J Pain* 2004;5:133–7.
11 Garde S, Knaup P, Hovenga E, *et al*. Towards semantic interoperability for electronic health records. *Methods Inf Med* 2007;46:332–43.

12 Olson S, Downey AS. Sharing Clinical Research Data: Workshop Summary. The National Academies Press, 2013. http://www.nap.edu/openbook.php?record_id=18267

13 Institute of Medicine. Data Harmonization for Patient-Centered Clinical Research – A Workshop. Washington, DC: Institute of Medicine, 2013. http://www.iom.edu/Activities/Quality/VSRT/~/media/Files/Activity%20Files/Quality/VSRT/Data-Harmonization/VSRT-WIB-DataHarmonization.pdf (accessed 29 Jan 2014).

14 Gardner SP. Ontologies and semantic data integration. *Drug Discov Today* 2005;10:1001–7.

15 Kunapareddy N, Mirhaji P, Richards D, *et al*. Information integration from heterogeneous data sources: a semantic web approach. *AMIA Annu Symp Proc* 2006;2006:992.

16 Sinaci AA, Laleci Erturkmen GB. A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains. *J Biomed Inform* 2013;46:784–94.

17 Office of the National Coordinator. Meaningful Use Stage 2. 2013. http://www.healthit.gov/policy-researchers-implementers/meaningful-use-stage-2 (accessed 15 Feb 2014).

18 Meaningful Use Regulations. http://www.healthit.gov/policy-researchers-implementers/meaningful-use-regulations (accessed 12 Feb 2014).

19 Centers for Medicare & Medicaid Services. EHR Incentive Programs: Stage 2. 2013. http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Stage_2.html (accessed 15 Feb 2014).

20 Toh S, Platt R, Steiner JF, *et al*. Comparative-effectiveness research in distributed health data networks. *Clin Pharmacol Ther* 2011;90:883–7.

21 Brown JS, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. *Med Care* 2013;51:S22–9.

22 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. *Med Care* 2012;50(Suppl):S21–9.

23 Kahn MG. Data Quality Collaborative. Data Quality Collaborative. 2012. http://repository.academyhealth.org/dqc/ (accessed 15 Feb 2014).

24 De Moor JS, Mariotto AB, Parry C, *et al*. Cancer survivors in the United States: prevalence across the survivorship trajectory and implications for care. *Cancer Epidemiol Biomarkers Prev* 2013;22:561–70.

25 Oster ME, Riehle-Colarusso T, Simeone RM, *et al*. Public health science agenda for congenital heart defects: report from a Centers for Disease Control and Prevention experts meeting. *J Am Heart Assoc* 2013;2:e000256.

26 CDC. 2011 National Diabetes Fact Sheet—Publications—Diabetes DDT. http://www.cdc.gov/diabetes/pubs/factsheet11.htm (accessed 12 Feb 2014).