hmo research network

HMORN: Infrastructure Bibliography and Website Materials

<u>Date of Last Update</u>: 7/6/2015

<u>Topic Area</u>:

I. General Description

1. Davis RL, Black S, Vadheim C, Shinefield H, Baker B, Pearson D, Chen R. Immunization tracking systems: experience of the CDC Vaccine Safety Datalink sites. *HMO Pract* 1997;11:13-17. PMID: 10165549.

    Recognizing the potential of large databases within HMOs for the evaluation of vaccine safety, the Centers for Disease Control and Prevention (CDC) funded the Vaccine Safety Datalink project, linking outcome and vaccine exposure information at Group Health Cooperative of Puget Sound, Kaiser Permanente Northwest, Kaiser Permanente Medical Care Program Northern California, and Southern California Kaiser Permanente. Integral to the Vaccine Safety Datalink Project was the development of Immunization Tracking Systems at each site; this report describes the effort required to establish these tracking systems. Essential requirements are the methods used to insure data quality and to educate system users. Tracking systems can be a valuable means for assessing vaccine coverage, evaluating barriers to complete immunization, and studying the effectiveness of interventions design to improve immunization coverage. Finally, we report on recent efforts to link HMO Immunization Tracking Systems with developing regional tracking systems.

2. Selby JV. Linking automated databases for research in managed care settings. Ann Intern Med. 1997;127(8 pt 2):719–724. PMID: 9382386.

    Spurred by demands for data from employer-purchasers and accreditation agencies and the adoption of strategies for disease management and outcome-based quality assurance, managed care organizations have recognized the need for rapid, convenient access to clinical information. Large investments in administrative and clinical data systems have also produced unprecedented opportunities for research on health care and epidemiology in large, defined populations. There is a long history of contributions to research by investigators who are based in the older nonprofit group and staff models of health maintenance organizations (HMOs). Many of these organizations maintain research units that are primarily funded by outside sources. Research includes descriptive and etiologic studies of epidemiology, randomized and observational studies of the effectiveness of treatment regimens, studies of disease costs and estimation of cost-effectiveness, investigations of risk predictions in populations, of risk and changes in organizational behavior, and evaluations of interventions to alter physician and patient behavior. The work is often conducted in collaboration with academic researchers. The HMO Research Network has recently been established to foster a scientific exchange among HMO-based researchers. As managed care organizations come to provide health care coverage to most U.S. citizens, research conducted by these organizations increasingly overlaps with public health research. Collaboration between HMO-based research centers and researchers from academia and government will undoubtedly continue to increase.

3. Chen RT, Glasser JW, Rhodes PH, Davis RL, Barlow WE, Thompson RS, Mullooly JP, Black SB, Shinefield HR, Vadheim CM, Marcy SM, Ward JI, Wise RP, Wassilak SG, Hadler SC. Vaccine Safety Datalink project: a new tool for improving vaccine safety monitoring in the United States. The Vaccine Safety Datalink Team. Pediatrics. 1997 Jun;99(6):765-73. PMID: 9164767.

Objective: To fill the large "gaps and limitations" in current scientific knowledge of rare vaccine adverse events identified in recent reviews of the Institute of Medicine.
Methods: Computerized information on immunization, medical outcomes, and potential confounders on more than 500 000 children 0 to 6 years of age is linked annually at several health maintenance organizations to create a large cohort for multiple epidemiologic studies of vaccine safety.
Results: Analysis of 3 years of follow-up data shows that 549 488 doses of diphtheria-tetanus-pertussis (DTP) and 310 618 doses of measles-mumps-rubella (MMR) vaccines have been administered to children in the study cohort. Analyses for associations between vaccines and 34 medical outcomes are underway. Screening of automated data shows that seizures are associated with receipt of DTP on the same day (relative risk [RR], 2.1; 95% confidence interval [CI], 1.1 to 4.0) and 8 to 14 days after receipt of MMR (RR, 3.0; 95% CI, 2.1 to 4.2). The diversity of vaccination exposures in this large cohort permits us to show that an apparent association of seizures 8 to 14 days after Haemophilus influenzae type b vaccine (RR, 1.6; 95% CI, 1.2 to 2.1) was attributable to confounding by simultaneous MMR vaccination; the association disappears with appropriate adjustment (RR, 1.0; 95% CI, 0.7 to 1.4).
Conclusion: Preliminary design, data collection, and analytic capability of the Vaccine Safety Datalink project has been validated by replication of previous known associations between seizures and DTP and MMR vaccines. The diversity in vaccine administration schedules permits potential disentangling of effects of simultaneous and combined vaccinations. The project provides a model of public health-managed care collaborations in addition to an excellent infrastructure for safety and other studies of vaccines.

4. Durham ML. Partnerships for Research among Managed Care Organizations. *Health AFF (Millwood).* 1998;17(1):111-122. PMID: 9455021.

   Researchers within health maintenance organizations (HMOs) need to create greater opportunities for collaborative research within their organizations. Multisite research will yield high-quality information for improving care. This paper describes situations in which competition as well as collaboration are possible across HMOs in the current environment. The paper considers the following questions: (1) What criteria determine if a project can be conducted as a multisite study; and (2) what population and organizational features should be considered when designing cross-site collaboration? The paper also discusses two important trends in the larger health care environment: cost containment, which is both a challenge and an opportunity for health services researchers working within managed care; and mergers and acquisitions, which are changing the face of the larger health care industry.

5. Chen RT, DeStefano F, Davis RL, Jackson LA, Thompson RS, Mullooly JP, Black SB, Shinefield HR, Vadheim CM., et al. The Vaccine Safety Datalink: immunization research in health maintenance organizations in the USA. *Bull World Health Organ* 2000;78:186-194. PMID: 10743283.

   The Vaccine Safety Datalink is a collaborative project involving the National Immunization Program of the Centers for Disease Control and Prevention and several large health maintenance organizations in the USA. The project began in 1990 with the primary purpose of rigorously evaluating concerns about the safety of vaccines. Computerized data on vaccination, medical outcome (e.g., outpatient visits, emergency room visits, hospitalizations, and deaths) and covariates (e.g. birth certificates, census data) are prospectively collected and linked under joint protocol

at multiple health maintenance organizations for analysis. Approximately 6 million persons (2% of the population of the USA) are now members of health maintenance organizations participating in the Vaccine Safety Datalink, which has proved to be a valuable resource providing important information on a number of vaccine safety issues. The databases and infrastructure created for the Vaccine Safety Datalink have also provided opportunities to address vaccination coverage, cost-effectiveness and other matters connected with immunization as well as matters outside this field.

6. 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;10:373–377. PMID: 11802579.

   Research and education programs in therapeutics that combine the data, organizational capabilities, and expertise of several managed care organizations working in concert can serve an important role when a single organization is not large enough to address a question of interest, when diversity in populations or delivery systems is required, and when it is necessary to establish consistency of results in different settings. Nine members of the HMO Research Network, a consortium of health maintenance organizations (HMOs) that perform public domain research, have formed a Center for Education and Research on Therapeutics (CERT), sponsored by the Agency for Healthcare Research and Quality, to conduct multicenter research in therapeutics. The CERT uses a distributed organizational model with shared leadership, in which data reside at the originating organization until they are needed to support a specific study. Extraction of data from the host computer systems, and some manipulation of data, is typically accomplished through computer programs that are developed centrally, then modified for use at each site. For complex studies, pooled analysis files are created by a coordinating center, and then analysed by investigators throughout the HMOs. It is also possible to contact HMO members when necessary. This multicenter environment has several benefits, addressing: (1) a wide array of questions about the safety and effectiveness of therapeutics, (2) the impact of efforts to change clinicians' and patients' behavior, and (3) pharmacoeconomic and pharmacogenetic questions.

7. DeStefano F, for the Vaccine Safety Datalink Research Group. The Vaccine Safety Datalink Project. Pharmacoepidemiology and Drug Safety 2001; 10: 403-406. PMID: 11802585.

   Summary: *The Vaccine Safety Datalink (VSD) is a collaborative project between the National Immunization Program of the Centers for Disease Control and Prevention (CDC) and several large health maintenance organizations (HMOs) in the United States. The project began in 1990 with the primary purpose of rigorously evaluating concerns about the safety of vaccines. Computerized data on vaccination, medical outcome (e.g. hospital discharge, outpatient visits, emergency room visits, and deaths), and covariate data (e.g. birth certificates and census) are prospectively collected at multiple HMOs (initially four) and linked under joint protocol for analyses. Approximately 6 million people (2% of the US population) are members of HMOs participating in the VSD. The VSD has proven to be a valuable resource that has provided important information on a number of vaccine safety issues. The databases and infrastructure created for the VSD have also provided opportunities to address other immunization questions including vaccination coverage and cost-effectiveness. In a recent investigation of intussusception following rotavirus vaccination, the VSD*

methodology was expanded to include 10 managed care organizations. A cohort study was conducted that allowed estimation of incidence rates of intussusception and attributable risks associated with rotavirus vaccine.

8. Wagner EH, Greene SM, Hart G, Field TS, Fletcher S, Geiger AM, Herrinton LJ, Hornbrook MC, Johnson CC, Mouchawar J, Rolnick SJ, Stevens VJ, Taplin SH, Tolsma D, Vogt TM. Building a research consortium of large health systems: the Cancer Research Network. J Natl Cancer Inst Monogr. 2005;(35):3-11. PMID: 16287880.

   Critical questions about cancer prevention, care, and outcomes increasingly require research involving large patient populations and their care delivery organizations. The Cancer Research Network (CRN) includes 11 integrated health systems funded by the National Cancer Institute (NCI) to conduct collaborative cancer research. This article describes the challenges of constructing a productive consortium of large health systems, and explores the CRN's responses. The CRN was initially funded through an NCI cooperative agreement in 1999 and has since received a second 4-year grant. Leadership and policy development are provided through a steering committee, subcommittees, and an external advisory committee. The CRN includes integral and affiliated research projects supported by a Scientific and Data Resources Core. Three characteristics of the CRN intensified the general challenges of consortium research: 1) its members are large health systems with legitimate concerns about confidentiality of data about enrollees, providers, and the organization; 2) CRN research projects often generate highly sensitive data about quality of care; and therefore 3) each participating organization wants a strong voice in CRN direction. CRN experience to date confirms that a consortium of health systems with internal research capacity can address a range of important cancer research questions that would be difficult to study in other venues. The advantages and challenges of consortium research are explored, with suggestions for the development, execution, and management of multisystem population laboratories.

9. Wallace PJ. Reshaping cancer learning through the use of health information technology. Health Aff (Millwood). 2007 Mar-Apr;26(2):w169-77. PMID: 17259200.

   In this paper I describe Kaiser Permanente's experience with health information technology (IT) in cancer care. Health IT holds the potential to accelerate learning in cancer care by comprehensively capturing rich patient data and supporting optimally standardized care. Rapid learning in cancer depends on simultaneously working toward universal technical and data standards and taking intermittent steps to reconcile variations in legacy systems through data-mapping and information-sharing initiatives.

10. Go AS, Magid DJ, Wells B, et al. The Cardiovascular Research Network: a new paradigm for cardiovascular quality and outcomes research. *Circ Cardiovasc Qual Outcomes.* 2008;1:138–147. PMID: 20031802.

    Background: *A clear need exists for a more systematic understanding of the epidemiology, diagnosis, and management of cardiovascular diseases. More robust data are also needed on how well clinical trials are translated into contemporary community practice and the associated resource use, costs, and outcomes.* ***Methods and Results***: *The National Heart, Lung, and Blood Institute recently established the Cardiovascular Research Network, which represents a new paradigm to evaluate the epidemiology, quality of care, and outcomes of cardiovascular disease and to conduct future clinical trials using a community-based*

model. The network includes 15 geographically distributed health plans with dedicated research centers, National Heart, Lung, and Blood Institute representatives, and an external collaboration and advisory committee. Cardiovascular research network sites bring complementary content and methodological expertise and a diverse population of 11 million individuals treated through various health care delivery models. Each site's rich electronic databases (e.g., sociodemographic characteristics, inpatient and outpatient diagnoses and procedures, pharmacy, laboratory, and cost data) are being mapped to create a standardized virtual data warehouse to facilitate rapid and efficient large-scale research studies. Initial projects focus on (1) hypertension recognition and management, (2) quality and outcomes of warfarin therapy, and (3) use, outcomes, and costs of implantable cardioverter defibrillators.

Conclusions*: The Cardiovascular Research Network represents a new paradigm in the approach to cardiovascular quality of care and outcomes research among community-based populations. Its unique ability to characterize longitudinally large, diverse populations will yield novel insights into contemporary disease and risk factor surveillance, management, outcomes, and costs. The Cardiovascular Research Network aims to become the national research partner of choice for efforts to improve the prevention, diagnosis, treatment, and outcomes of cardiovascular diseases.*

11. Magid, DJ, Gurwitz, JH; Rumsfeld, JS; Go, AS. Creating a research data network for cardiovascular disease: the CVRN. *Expert Rev Cardiovasc Ther*. 2008;6:1043–1045. PMID: 18793105. **Editorial - no abstract.**

12. Geiger AM, Buist DS, Greene SM, Altschuler A, Field TS. Survivorship research based in integrated healthcare delivery systems: the Cancer Research Network. Cancer 2008 06.01;112(11 Suppl):2617-26. PMID: 18428194.

Background: *Integrated healthcare delivery systems present unique opportunities for cancer survivorship research. The National Cancer Institute funds the Cancer Research Network (CRN) to leverage these capabilities for all types of cancer research, including survivorship.*

Methods: *The authors gathered information from a recent CRN funding application, Survivorship Interest Group materials, the CRN website, and published articles. CRN studies were selected to illustrate diverse topics and a variety of data-collection approaches.*

Results: *The 14 systems that participate in the CRN provide care for approximately 10.8 million individuals of all ages and racial/ethnic backgrounds, for whom approximately 38,000 new cancer diagnoses were made in 2005. CRN systems have the ability to use existing data and collect new data on patients, providers, and organizations through well-established research centers staffed by independent scientists. Of the 45 funded and 2 pending CRN grant applications as of November 30, 2007, 21 include aspects related to cancer survivorship. These studies have examined clinical trial participation, patterns of care, age and racial/ethnic disparities, diffusion of clinical trial findings, treatment outcomes, surveillance, and end-of-life and palliative care. Breast, colorectal, lung, ovarian, and prostate cancers have been the focus of these studies. Results of these studies have been published widely in leading journals.*

Conclusions: *Completed and ongoing CRN survivorship studies provide a strong foundation for future studies. Scientists from all institutional affiliations are welcome*

to approach the CRN with ideas and are encouraged to allow ample time to establish collaborative relationships and design rigorous studies.
(intro paragraph describes HMORN).

13. Davis RL. Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP). J Popul Ther Clin Pharmacol. 2010 Fall;17(3):e336-40. PMID: 21041866.

Knowledge about safe medication use during pregnancy is limited, yet about two of every three women take at least one prescription medication during pregnancy, furthermore, there is a lack of rigorous studies evaluating birth outcomes associated with in utero exposure to medications. The Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP) is intended to provide a mechanism for collaborative pharmacoepidemiological research to address the safety of pharmaceutical product exposure during pregnancy, through the development of standardized data requirements and of the necessary data linkages of mother-infant pairs to conduct multi-site investigations. This presentation will describe the program, the types of data collected, and progress to date. The current MEPREP population includes female health plan members of 11 distinct health management entities within three research centres who have delivered an infant between January 1, 2001 and December 31, 2007, along with the administrative and birth certificate data on over one million children linked to mothers. There is information on all the medications those mothers took, as well as most of the outcomes of the babies. One of the benefits of this dataset is the information that could be investigated, such as birth weight, fetal growth, congenital anomalies, perinatal conditions, etc., against various demographics of the women in the dataset. The population size within the dataset suggests that various parameters could be studied with at least a modest degree of power.

14. Baggs J, Gee J, Lewis E, Fowler G, Benson P, Lieu T, Naleway A, Klein NP, Baxter R., et al. The Vaccine Safety Datalink: a model for monitoring immunization safety. *Pediatrics* 2011;127 Suppl 1:S45-S53. PMID: 21502240.

The Vaccine Safety Datalink (VSD) project is a collaborative project between the Centers for Disease Control and Prevention and 8 managed care organizations (MCOs) in the United States. Established in 1990 to conduct postmarketing evaluations of vaccine safety, the project has created an infrastructure that allows for high-quality research and surveillance. The 8 participating MCOs comprise a large population of 8.8 million members annually (3% of the US population), which enables researchers to conduct studies that assess adverse events after immunization. Each MCO prepares computerized data files by using a standardized data dictionary containing demographic and medical information on its members, such as age and gender, health plan enrollment, vaccinations, hospitalizations, outpatient clinic visits, emergency department visits, urgent care visits, and mortality data, as well as additional birth information (eg, birth weight) when available. Other information sources, such as medical chart review, member surveys, and pharmacy, laboratory, and radiology data, are often used in VSD studies to validate outcomes and vaccination data. Since 2000, the VSD has undergone significant changes including an increase in the number of participating MCOs and enrolled population, changes in data-collection procedures, the creation of near real-time data files, and the development of near real-time postmarketing surveillance for newly licensed vaccines or changes in vaccine recommendations. Recognized as an important resource in vaccine safety, the VSD is working toward increasing transparency

through data-sharing and external input. With its recent enhancements, the VSD provides scientific expertise, continues to develop innovative approaches for vaccine-safety research, and may serve as a model for other patient safety collaborative research projects.

15. Andrade SE, Davis RL, Cheetham TC, Cooper WO, Li DK, Amini T, Beaton SJ, Dublin S, Hammad TA, Pawloski PA, Raebel MA, Smith DH, Staffa JA, Toh S, Dashevsky I, Haffenreffer K, Lane K, Platt R, Scott PE. Medication Exposure in Pregnancy Risk Evaluation Program. Matern Child Health J. 2012 Oct;16(7):1349-54. PMID: 22002179.

To describe a program to study medication safety in pregnancy, the Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP). MEPREP is a multi-site collaborative research program developed to enable the conduct of studies of medication use and outcomes in pregnancy. Collaborators include the U.S. Food and Drug Administration and researchers at the HMO Research Network, Kaiser Permanente Northern and Southern California, and Vanderbilt University. Datasets have been created at each site linking healthcare data for women delivering an infant between January 1, 2001 and December 31, 2008 and infants born to these women. Standardized data files include maternal and infant characteristics, medication use, and medical care at 11 health plans within 9 states; birth certificate data were obtained from the state departments of public health. MEPREP currently involves more than 20 medication safety researchers and includes data for 1,221,156 children delivered to 933,917 mothers. Current studies include evaluations of the prevalence and patterns of use of specific medications and a validation study of data elements in the administrative and birth certificate data files. MEPREP can support multiple studies by providing information on a large, ethnically and geographically diverse population. This partnership combines clinical and research expertise and data resources to enable the evaluation of outcomes associated with medication use during pregnancy.

16. Andrade SE, Raebel MA, Boudreau D, Davis RL, Haffenreffer K, Pawloski PA, Toh S, Platt R. Health Maintenance Organizations/Health Plans. In: Eds: Strom BL, Kimmel SE, Hennessy S. *Pharmacoepidemiology*. Chapter 12, pp 163-188, John Wiley & Sons, Chichester, West Sussex, UK. 2012. **Book chapter - no abstract.**

II.   Relationships with Sponsoring Health Plan/Organization

17. Vogt TM, Elston-Lafata J, Tolsma D, Greene SM. The role of research in integrated healthcare systems: the HMO Research Network. Am J Manag Care 2004; 10:643-8. PMID: 15515997.

Science is the basis of medicine. Good science leads to better decisions and more effective systems to support those decisions. Most individuals associate science primarily with academic institutions. However, top-quality research relevant to managing the health of populations and the care of specific clinical conditions is increasingly being carried out by investigators working in integrated healthcare systems. This introduction outlines the activities of the HMO Research Network, whose researchers have made and continue to make important contributions to the field of health research. Its objective is to inform readers of the activities and value of systems-based health research. We describe the importance and extent of the

research conducted by HMO Research Network members, as well as the advantages of conducting research in such settings.

18. Somkin CP, Altschuler A, Ackerson L, Geiger AM, Greene SM, Mouchawar J, Holup J, Fehrenbacher L, Nelson A, Glass A, Polikoff J, Tishler S, Schmidt C, Field T, Wagner E. Organizational barriers to physician participation in cancer clinical trials. Am J Manag Care. 2005 Jul;11(7):413-21. PMID: 16044978.

Objective: To assess barriers to physician participation in cancer clinical trials among oncologists, oncology leaders, and health plan leaders
Study Design: Mail survey of 221 oncologists combined with semistructured telephone interviews with oncology and plan leaders at 10 integrated healthcare systems.
Methods: The survey instrument examined physicians' involvement in clinical trials; their perception of the value of trials to them, their patients, and their organization; and the presence of infrastructure support for trials and associated resource constraints. The interviews investigated similar issues from the leaders' perspective. We used linear regression to model trial enrollment and standard qualitative techniques to analyze the interviews.
Results: Oncologists estimated they enrolled 7% of patients in trials. They expressed extremely favorable attitudes toward trials as a source of high-quality patient care and a benefit to themselves professionally. While positive attitudes toward trials were common, and were significant bivariate predictors of enrollment, organizational factors were the predominant predictors in multivariate analysis. The best combination of factors independently predicting enrollment related to organizational support for trials, subspecialty of the oncologist, and limitations of trial eligibility requirements.
Conclusions: To increase trial participation, there is a critical need for infrastructure to support trials, especially additional support staff and research nurses. In addition, there is a need for better intra-organizational communication and consideration of the impact of trial design on internal health plan resources. This research supports the need to continue a national dialogue about the broadly defined benefits and costs of clinical trials to patients, physicians, and health plans.

19. Somkin CP, Altschuler A, Ackerson L, Tolsma D, Rolnick SJ, Yood R, Weaver WD, Von Worley A, Hornbrook M, Magid DJ, Go AS. Cardiology clinical trial participation in community-based healthcare systems: obstacles and opportunities. Contemp Clin Trials. 2008 Sep;29(5):646-53. PMID: 18397842.

Background: The objective of our study was to examine cardiologists' and organizational leaders' interest in clinical trial participation and perceived barriers and facilitators to participation within ten diverse non-profit healthcare delivery systems. Trials play a pivotal role in advancing knowledge about the safety and efficacy of cardiovascular interventions and tests. Although cardiovascular trials successfully enroll patients, recruitment challenges persist. Community-based health systems could be an important source of participants and investigators, but little is known about community cardiologists' experiences with trials.
Methods: We interviewed 25 cardiology and administrative leaders and mailed questionnaires to all 280 cardiologists at 10 U.S. healthcare organizations.
Results: The survey received a 73% response rate. While 60% of respondents had

not participated in any trials in the past year, nearly 75% wanted greater participation. Cardiologists reported positive attitudes toward trial participation; more than half agreed that trials were their first choice of therapy for patients, if available. Almost all leaders described their organizations as valuing research but not necessarily trials. Major barriers to participation were lack of physician time and insufficient skilled research nurses.

Conclusions*: Cardiologists have considerable interest in trial participation. Major obstacles to increased participation are lack of time and effective infrastructure to support trials. These results suggest that community-based health systems are a rich source for cardiovascular research but additional funding and infrastructure are needed to leverage this resource.*

20. Mazor KM, Sabin JE, Goff SL, Smith DH, Rolnick S, Roblin D, Raebel MA, Herrinton LJ, Gurwitz JH, Boudreau D, Meterko V, Dodd KS, Platt R. Cluster randomized trials to study the comparative effectiveness of therapeutics: stakeholders' concerns and recommendations. Pharmacoepidemiol Drug Saf. 2009 Jul;18(7):554-61. PMID: 19402030.

    Purpose: *To describe the concerns raised by health plan members, providers and purchasers related to studying the comparative effectiveness of therapeutics using cluster randomized trials (CRTs) within health plans. An additional goal was to develop recommendations for increasing acceptability.*
    Methods: *Eighty-four qualitative in-depth telephone interviews were conducted; 50 with health plan members, 21 with providers, and 13 with purchasers. Interviews focused on stakeholders' concerns about and recommendations for conducting CRTs in health plans.*
    Results: *Members expressed concerns that CRTs might compromise their healthcare. Providers and purchasers recognized the value of and the need for comparative effectiveness research. Providers expressed concerns that they would not have sufficient time to discuss a CRT with patients, and that participation in such a study could negatively impact their relationships with patients. Purchasers would want assurances that study participation would not result in members receiving lesser care, and that benefits would remain equitable for all members.*
    Conclusions: *This study provides insight into how health plan members, providers and purchasers might react to a CRT being conducted in their health plan. The recommendations reported here provide guidance for researchers and policy makers considering this methodological approach and suggest that with sufficient preparation and planning CRTs can be an acceptable and efficient methodology for studying the comparative effectiveness of therapeutics in real world settings.*

21. Schmittdiel JA, Grumbach, K, Selby JV. System-based Participatory Research in Health Care: An approach for Sustainable Translational Research and Quality Improvement. Ann Fam Med 2010;May/June 8(3): 256-259. PMID: 20458110.

    Translational research seeks to improve health care by promoting action and change in real-world health care settings. Although translational research advocates a break from the traditional researcher-initiated approach to science, strategies to successfully engage clinicians and leaders of health care delivery organizations in research are still under development. We propose that applying the principles of community-based participatory research in a way that considers delivery systems— including their leaders, clinicians, and staff—as a focal community can enhance the ability of translational research to improve health care. Applying participatory

research methods, such as engaging in collaborative partnerships, building on existing community strengths, investing in long-term relationships, and engaging in research as a cyclical, iterative process, can be a successful approach to sustainable quality improvement at the systems level.

22. Greene SM, Reid RJ, Larson EB. Implementing the Learning Health System: from Concept to Action. Ann Intern Med 2012; 157: 207-210. PMID: 22868839.

Clinicians and health systems are facing widespread challenges, including changes in care delivery, escalating health care costs, and the need to keep up with rapid scientific discovery. Reorganizing U.S. health care and changing its practices to render better, more affordable care requires transformation in how health systems generate and apply knowledge. The "rapid-learning health system"-posited as a conceptual strategy to spur such transformation-leverages recent developments in health information technology and a growing health data infrastructure to access and apply evidence in real time, while simultaneously drawing knowledge from real-world care-delivery processes to promote innovation and health system change on the basis of rigorous research. This article describes an evolving learning health system at Group Health Cooperative, the 6 phases characterizing its approach, and examples of organization-wide applications. This practical model promotes bidirectional discovery and an open mind at the system level, resulting in willingness to make changes on the basis of evidence that is both scientifically sound and practice-based. Rapid learning must be valued as a health system property to realize its full potential for knowledge generation and application.

23. Lowry SJ, Loggers ET, Bowles EJ, Wagner EH. Evidence gaps in advanced cancer care: community-based clinicians' perspectives and priorities for comparative effectiveness research. *J Oncol Pract* 2012;8:28s-33s. PMID: 22942821.

***Purpose:** Although much effort has focused on identifying national comparative effectiveness research (CER) priorities, little is known about the CER priorities of community-based practitioners treating patients with advanced cancer. CER priorities of managed-care-based clinicians may be valuable as reflections of both payer and provider research interests.*
***Methods:** We conducted mixed methods interviews with 10 clinicians (five oncologists and five pharmacists) at five health plans within the Health Maintenance Organization Cancer Research Network. We asked, "What evidence do you most wish you had when treating patients with advanced cancer?" and questioned participants on their impressions and knowledge of CER and pragmatic clinical trials (PCTs). We conducted qualitative analyses to identify themes across interviews.*
***Results:** Ninety percent of participants had heard of CER, 20% had heard of PCTs, and all rated CER/PCTs as highly relevant to patient and health plan decision making. Each participant offered between three and 10 research priorities. Half (49%) involved head-to-head treatment comparisons; another 20% involved comparing different schedules or dosing regimens of the same treatment. The majority included alternative outcomes to survival (e.g., toxicity, quality of life, non-inferiority). Participants cited several limitations to existing evidence including lack of generalizability, funding biases, and rapid development of new treatments.*
***Conclusion:** Head-to-head treatment comparisons remain a major evidence need among community-based oncology clinicians, and CER/PCTs are highly valued methods to address the limitations of traditional randomized trials, answer questions*

of cost-effectiveness or non-inferiority, and inform data-driven dialogue and decision making by all stakeholders.

III.   Informatics/Distributed Research Models/Data Models (VDW)

24. Platt R, Bocchino C, Caldwell B, Harmon R, Kleinman K, Lazarus R, Nelson AF, Nordin JD, Ritzwoller DP. Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program. J Urban Health. 2003 Jun;80(2 Suppl 1):i25-31. PMID: 12791776.

Several health plans and other organizations are collaborating with the Centers for Disease Control and Prevention to develop a syndromic surveillance system with national coverage that includes more than 20 million people. A principal design feature of this system is reliance on daily reporting of counts of individuals with syndromes of interest in specified geographic regions rather than reporting of individual encounter-level information. On request from public health agencies, health plans and telephone triage services provide additional information regarding individuals who are part of apparent clusters of illness. This reporting framework has several advantages, including less sharing of protected health information, less risk that confidential information will be distributed inappropriately, the prospect of better public acceptance, greater acceptance by health plans, and less effort and cost for both health plans and public health agencies. If successful, this system will allow any organization with appropriate data to contribute vital information to public health syndromic surveillance systems while preserving individuals' privacy to the greatest extent possible.

25. Hornbrook MC, Hart G, Ellis JL, Bachman DJ, Ansell G, Greene SM, Wagner EH, Pardee R, Schmidt MM, Geiger A, Butani AL, Field T, Fouayzi H, Miroshnik I, Liu L, Diseker R, Wells K, Krajenta R, Lamerato L, Neslund Dudas C. Building a virtual cancer research organization. J Natl Cancer Inst Monogr 2005(35):12-25. PMID: 16287881.

Background: *The Cancer Research Network (CRN) comprises the National Cancer Institute and 11 non-profit research centers affiliated with integrated health care delivery systems. The CRN, a public/private partnership, fosters multisite collaborative research on cancer prevention, screening, treatment, survival, and palliation in diverse populations.*
Methods: *The CRN's success hinges on producing innovative cancer research that likely would not have been developed by scientists working individually, and then translating those findings into clinical practice within multiple population laboratories. The CRN is a collaborative virtual research organization characterized by user-defined sharing among scientists and health care providers of data fi les as well as direct access to researchers, computers, software, data, research participants, and other resources. The CRN's research management Web site fosters a high-functioning virtual scientific community by publishing standardized data definitions, file specifications, and computer programs to support merging and analyzing data from multiple health care systems.*
Results: *Seven major types of standardized data files developed to date include demographics, health plan eligibility, tumor registry, inpatient and ambulatory utilization, medication dispensing, laboratory tests, and imaging procedures; more will follow. Data standardization avoids rework, increases multisite data integrity,*

increases data security, generates shorter times from initial proposal concept to submission, and stimulates more frequent collaborations among scientists across multiple institutions. Conclusions: The CRN research management Web site and associated standardized data fi les and procedures represent a quasi-public resource, and the CRN stands ready to collaborate with researchers from outside institutions in developing and conducting innovative public domain research.

26. Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. BMC Public Health. 2006 Sep 19;6:235. PMID: 16984658.

    Background: *Many systems for routine public health surveillance rely on centralized collection of potentially identifiable, individual, identifiable personal health information (PHI) records. Although individual, identifiable patient records are essential for conditions for which there is mandated reporting, such as tuberculosis or sexually transmitted diseases, they are not routinely required for effective syndromic surveillance. Public concern about the routine collection of large quantities of PHI to support non-traditional public health functions may make alternative surveillance methods that do not rely on centralized identifiable PHI databases increasingly desirable.*
    Methods: *The National Bioterrorism Syndromic Surveillance Demonstration Program (NDP) is an example of one alternative model. All PHI in this system is initially processed within the secured infrastructure of the health care provider that collects and holds the data, using uniform software distributed and supported by the NDP. Only highly aggregated count data is transferred to the datacenter for statistical processing and display.*
    Results: *Detailed, patient level information is readily available to the health care provider to elucidate signals observed in the aggregated data, or for ad hoc queries. We briefly describe the benefits and disadvantages associated with this distributed processing model for routine automated syndromic surveillance.*
    Conclusion: *For well-defined surveillance requirements, the model can be successfully deployed with very low risk of inadvertent disclosure of PHI--a feature that may make participation in surveillance systems more feasible for organizations and more appealing to the individuals whose PHI they hold. It is possible to design and implement distributed systems to support non-routine public health needs if required.*

27. Maro JC, Platt R, Holmes JH, Strom BL, Hennessy S, Lazarus R, Brown JS. Design of a National Distributed Health Data Network. Ann Intern Med. 2009;151:341-344. PMID: 19638403.

    A distributed health data network is a system that allows secure remote analysis of separate data sets, each comprising a different medical organization's or health plan's records. Distributed health data networks are currently being planned that could cover millions of people, permitting studies of comparative clinical effectiveness, best practices, diffusion of medical technologies, and quality of care. These networks could also support assessment of medical product safety and other public health needs. Distributed network technologies allow data holders to control all uses of their data, which overcomes many practical obstacles related to confidentiality, regulation, and proprietary interests. Some of the challenges and potential methods of operation of a multipurpose, multi-institutional distributed health data network are described.

28. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. 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–S51. PMID: 20473204.

Background: Comparative effectiveness research, medical product safety evaluation, and quality measurement will require the ability to use electronic health data held by multiple organizations. There is no consensus about whether to create regional or national combined (eg, "all payer") databases for these purposes, or distributed data networks that leave most Protected Health Information and proprietary data in the possession of the original data holders.
Objectives: Demonstrate functions of a distributed research network that supports research needs and also address data holders concerns about participation. Key design functions included strong local control of data uses and a centralized web-based querying interface.
Research Design: We implemented a pilot distributed research network and evaluated the design considerations, utility for research, and the acceptability to data holders of methods for menu-driven querying. We developed and tested a central, web-based interface with supporting network software. Specific functions assessed include query formation and distribution, query execution and review, and aggregation of results.
Results: This pilot successfully evaluated temporal trends in medication use and diagnoses at 5 separate sites, demonstrating some of the possibilities of using a distributed research network. The pilot demonstrated the potential utility of the design, which addressed the major concerns of both users and data holders. No serious obstacles were identified that would prevent development of a fully functional, scalable network.
Conclusions: Distributed networks are capable of addressing nearly all anticipated uses of routinely collected electronic healthcare data. Distributed networks would obviate the need for centralized databases, thus avoiding numerous obstacles.

29. Toh S, Platt R, Steiner JF, Brown JS. Comparative-Effectiveness Research in Distributed Health Data Networks. Clinical Pharmacology & Therapeutics. 2011 Dec;90(6):883-887. PMID: 22030567.

Comparative-effectiveness research (CER) can be conducted within a distributed health data network. Such networks allow secure access to separate data sets from different data partners and overcome many practical obstacles related to patient privacy, data security, and proprietary concerns. A scalable network architecture supports a wide range of CER activities and meets the data infrastructure needs envisioned by the Federal Coordinating Council for Comparative Effectiveness Research.

30. Sittig DF, Hazlehurst BL, Brown J, Murphy, S, Rosenman, M, Tarczy-Hornoch, P, Wilcox, AB. A Survey of Informatics Platforms That Enable Distributed Comparative Effectiveness Research Using Multi-institutional Heterogenous Clinical Data. Med Care 2012 Jul;50 Suppl: S49–S59. PMID: 22692259.

Comparative effectiveness research (CER) has the potential to transform the current health care delivery system by identifying the most effective medical and surgical treatments, diagnostic tests, disease prevention methods, and ways to deliver care for specific clinical conditions. To be successful, such research requires the identification, capture, aggregation, integration, and analysis of disparate data

sources held by different institutions with diverse representations of the relevant clinical events. In an effort to address these diverse demands, there have been multiple new designs and implementations of informatics platforms that provide access to electronic clinical data and the governance infrastructure required for inter-institutional CER. The goal of this manuscript is to help investigators understand why these informatics platforms are required and to compare and contrast 6 large-scale, recently funded, CER-focused informatics platform development efforts. We utilized an 8-dimension, sociotechnical model of health information technology to help guide our work. We identified 6 generic steps that are necessary in any distributed, multi-institutional CER project: data identification, extraction, modeling, aggregation, analysis, and dissemination. We expect that over the next several years these projects will provide answers to many important, and heretofore unanswerable, clinical research questions.

31. Curtis L, Weiner M, Boudreau DM, Cooper WO, Daniel GW, Nair VP, Raebel MA, Beaulieu NU, Rosofsky R, Woodworth TS, Brown JS. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidemiology and Drug Safety 2012; 21(S1): 23–31. PMID: 22262590

Purpose: *We describe the design, implementation, and use of a large, multiorganizational distributed database developed to support the Mini-Sentinel Pilot Program of the US Food and Drug Administration (FDA). As envisioned by the US FDA, this implementation will inform and facilitate the development of an active surveillance system for monitoring the safety of medical products (drugs, biologics, and devices) in the USA.*
Methods: *A common data model was designed to address the priorities of the Mini-Sentinel Pilot and to leverage the experience and data of participating organizations and data partners. A review of existing common data models informed the process. Each participating organization designed a process to extract, transform, and load its source data, applying the common data model to create the Mini-Sentinel Distributed Database. Transformed data were characterized and evaluated using a series of programs developed centrally and executed locally by participating organizations. A secure communications portal was designed to facilitate queries of the Mini-Sentinel Distributed Database and transfer of confidential data, analytic tools were developed to facilitate rapid response to common questions, and distributed querying software was implemented to facilitate rapid querying of summary data.*
Results: *As of July 2011, information on 99,260,976 health plan members was included in the Mini-Sentinel Distributed Database. The database includes 316,009,067 person-years of observation time, with members contributing, on average, 27.0 months of observation time. All data partners have successfully executed distributed code and returned findings to the Mini-Sentinel Operations Center.*
Conclusion: *This work demonstrates the feasibility of building a large, multiorganizational distributed data system in which organizations retain possession of their data that are used in an active surveillance system.*

32. Nichols GA, Desai J, Elston Lafata J, Lawrence JM, O'Connor PJ, Pathak RD, et al. Construction of a Multisite DataLink Using Electronic Health Records for the Identification, Surveillance, Prevention, and Management of Diabetes Mellitus: The SUPREME-DM Project. Prev Chronic Dis 2012;9:E110. PMID: 22677160.

Introduction: Electronic health record (EHR) data enhance opportunities for conducting surveillance of diabetes. The objective of this study was to identify the number of people with diabetes from a diabetes DataLink developed as part of the SUPREME-DM (SUrveillance, PREvention, and ManagEment of Diabetes Mellitus) project, a consortium of 11 integrated health systems that use comprehensive EHR data for research.
Methods: We identified all members of 11 health care systems who had any enrollment from January 2005 through December 2009. For these members, we searched inpatient and outpatient diagnosis codes, laboratory test results, and pharmaceutical dispensings from January 2000 through December 2009 to create indicator variables that could potentially identify a person with diabetes. Using this information, we estimated the number of people with diabetes and among them, the number of incident cases, defined as indication of diabetes after at least 2 years of continuous health system enrollment.
Results: The 11 health systems contributed 15,765,529 unique members, of whom 1,085,947 (6.9%) met 1 or more study criteria for diabetes. The nonstandardized proportion meeting study criteria for diabetes ranged from 4.2% to 12.4% across sites. Most members with diabetes (88%) met multiple criteria. Of the members with diabetes, 428,349 (39.4%) were incident cases.
Conclusion: The SUPREME-DM DataLink is a unique resource that provides an opportunity to conduct comparative effectiveness research, epidemiologic surveillance including longitudinal analyses, and population-based care management studies of people with diabetes. It also provides a useful data source for pragmatic clinical trials of prevention or treatment interventions.

33. America's Health Insurance Plans. The Vaccine Safety Datalink: An effective model for monitoring vaccine safety—Growth and accomplishments from 2002 to 2012. March 2013.

The Vaccine Safety Datalink (VSD) was created in 1990 by the Centers for Disease Control and Prevention (CDC) as a crucial part of the federal government's effort to monitor the safety of vaccines used in the United States and to reassure public confidence in vaccines. Through the VSD project, researchers from participating health plans and CDC have worked together on studies designed to monitor the safety of vaccines administered to infants, children, adolescents, and adults. Beginning as a system of four integrated health plans to study the safety of childhood immunizations, the VSD evolved and expanded gradually to become a valuable tool for the national immunization program. As the VSD entered its second decade, the CDC sought to significantly expand the scope of the VSD in order to address a wide array of safety issues, including new vaccines for both children and adults.

34. Ross TR, Ng D, Brown JS, Pardee R, Hornbrook M, Hart G, Steiner JF. The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. eGEMs (Generating Evidence & Methods to improve patient outcomes) 2014 Vol. 2: Iss. 1, Article 2. DOI: http://dx.doi.org/10.13063/2327-9214.1049 Available at: http://repository.academyhealth.org/egems/vol2/iss1/2

The HMO Research Network (HMORN) Virtual Data Warehouse (VDW) is a public, non-proprietary, research-focused data model implemented at 17 health care systems across the United States. The HMORN has created a governance structure and specified policies concerning the VDW's content, development, implementation, and quality assurance. Data extracted from the VDW have been used by thousands of studies published in peer-reviewed journal articles. Advances in software supporting care delivery and claims processing and the availability of new data sources have greatly expanded the data available for research, but substantially increased the complexity of data management. The VDW data model incorporates software and data advances to ensure that comprehensive, up-to-date data of known quality are available for research. VDW governance works to accommodate new data and system complexities. This article highlights the HMORN VDW data model, its governance principles, data content and quality assurance procedures. Our goal is to share the VDW data model and its operations to those wishing to implement a distributed interoperable health care data system.

35. Johnson, Karin E.; Kamineni, Aruna; Fuller, Sharon; Olmstead, Danielle; and Wernli, Karen J. (2014) "How the Provenance of Electronic Health Record Data Matters for Research: A Case Example Using System Mapping," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 2: Iss. 1, Article 4. DOI: http://dx.doi.org/10.13063/2327-9214.1058. Available at: http://repository.academyhealth.org/egems/vol2/iss1/4

Introduction: The use of electronic health records (EHRs) for research is proceeding rapidly, driven by computational power, analytical techniques, and policy. However, EHR-based research is limited by the complexity of EHR data and a lack of understanding about data provenance, meaning the context under which the data were collected. This paper presents system flow mapping as a method to help researchers more fully understand the provenance of their EHR data as it relates to local workflow. We provide two specific examples of how this method can improve data identification, documentation, and processing.

Background: EHRs store clinical and administrative data, often in unstructured fields. Each clinical system has a unique and dynamic workflow and EHR, which may be influenced by broader context such as documentation required for billing.

Methods: We present a case study with two examples of using system flow mapping to characterize EHR data for a local colorectal cancer screening process.

Findings: System flow mapping demonstrated that information entered into the EHR during clinical practice required interpretation and transformation before it could be

accurately applied to research. We illustrate how system flow mapping shaped our knowledge of the quality and completeness of data in two examples: 1) determining colonoscopy indication as recorded in the EHR, and 2) discovering a specific EHR form that captured family history.

Discussion: Researchers who do not consider data provenance risk compiling data that are systematically incomplete or incorrect. For example, researchers who are not familiar with the clinical workflow under which data were entered might miss or misunderstand patient information or procedure and diagnostic codes.

Conclusions/Next steps: Data provenance is a fundamental characteristic of research data from EHRs. Given the diversity of EHR platforms and system workflows, researchers need tools for evaluating and reporting data availability, quality, and transformations. Our case study illustrates how system mapping can inform researchers about the provenance of their data as it pertains to local workflows.

IV. Data Quality and Validity

36. Mullooly J, Drew L, DeStefano F, Chen R, Okoro K, Swint E, Immanuel V, Ray P, Lewis N., et al. Quality of HMO vaccination databases used to monitor childhood vaccine safety. Vaccine Safety DataLink Team. *Am J Epidemiol* 1999;149:186-194. PMID: 9921964.

    The availability of large, population-based, automated, medical care databases provides unique opportunities for monitoring the safety of childhood vaccines. The authors assessed the quality of automated vaccination databases by comparing them with vaccinations documented in paper-based medical records at three large US West Coast health maintenance organizations (HMOs) participating in the Vaccine Safety DataLink (VSD) study, a Centers for Disease Control and Prevention collaborative study of childhood vaccine safety. The authors randomly selected 1% or 2% samples of VSD study populations (n = 1,224-2,577) for data quality analyses. Agreement between automated and abstracted vaccinations required identical triads of child identification number, vaccination date, and vaccine type. Separate analyses were conducted for each HMO and for each vaccine type administered between 1991 and 1995. Agreement was measured by three matching proportions: 1) the proportion of automated vaccinations present in the abstracted source, 2) the proportion of abstracted vaccinations present in the automated source, and 3) the proportion of vaccinations from either source present in both sources. Overall, for common childhood vaccines, proportion 1 ranged from 83% to 99%, proportion 2 ranged from 82% to 98%, and proportion 3 ranged from 70% to 97%. Lack of automated data was the most frequent type of discrepancy, followed by date mismatches and vaccine type mismatches. Vaccination exposure classification errors in the range reported here were found by mathematical modeling to only modestly bias measured medical outcome rate ratios toward the null hypothesis. The results of the data quality analyses support the usefulness of vaccination exposure data derived from these automated HMO vaccination databases.

37. Ford ME, Hill DD, Nerez D, Hornbrook M, Zapka J, Meenan R, Greene S, Johnson CC. Categorizing race and ethnicity in the HMO Research Network. Ethnicity and Disease 2002;12:135-140.

Background: The Cancer Research Network (CRN) was formed in 1999 with funding from the National Cancer Institute. The CRN represents a collaboration of 10 health plans across the United States, with a combined total of approximately 9 million enrollees. The goal of the CRN is to promote collaborative research, which will ultimately increase the effectiveness of preventive, curative, and supportive interventions for major cancers. Special emphasis is placed upon diverse populations, and racial and ethnic differences in outcomes, costs, and cost effectiveness.
Purpose: There is increasing awareness in the research literature of the relationship between race and ethnicity and health outcomes. However, the majority of the health maintenance organizations represented in the CRN, similar to other health plans and organizations, do not routinely collect race and ethnicity data on their members. In order to compare data and outcomes across the CRN sites, consensus is needed in the measurement of race and ethnicity.
Methods: This review discusses terminology used in the research literature to describe race and ethnicity and the manner in which these constructs have been measured in previous studies.
Conclusions: This review concludes with suggestions for standardized measures of race and ethnicity.
Implications: It is hoped that shared conceptualizations of race and ethnicity will lead to improved data quality and precision in measurement.

38. Mullooly J, Drew L, DeStefano F, Maher J, Bohlke K, Immanuel V, Black S, Lewis E, Ray P., et al. Quality assessments of HMO diagnosis databases used to monitor childhood vaccine safety. *Methods Inf Med* 2004;43:163-170. PMID: 15136866.

Objective: To assess the quality of automated diagnoses extracted from medical care databases by the Vaccine Safety Datalink (VSD) study.
Methods: Two methods are used to assess quality of VSD diagnosis data. The first method compares common automated and abstracted diagnostic categories ("outcomes") in 1-2% simple random samples of study populations. The second method estimates positive predictive values of automated diagnosis codes used to identify potential cases of rare conditions (e.g., acute ataxia) for inclusion in nested case-control medical record abstraction studies.
Results: There was good agreement (64-68%) between automated and abstracted outcomes in the 1-2% simple random samples at 3 of the 4 VSD sites and poor agreement (44%) at 1 site. Overall at 3 sites, 56% of children with automated cerebella ataxia codes (ICD-9 = 334) and 22% with "lack of coordination" codes (ICD-9 = 781.3) met objective clinical criteria for acute ataxia.
Conclusions: The misclassification error rates for automated screening outcomes substantially reduce the power of screening analyses and limit usefulness of screening analyses to moderate to strong vaccine-outcome associations. Medical record verification of outcomes is needed for definitive assessments.

39. Hazlehurst B, Mullooly J, Naleway A, Crane B. Detecting possible vaccination reactions in clinical notes. *AMIA Annu Symp Proc* 2005;306-310. PMID: 16779051.

The Vaccine Safety Datalink is a collaboration between the CDC and eight large HMO's to investigate adverse events following immunization through analysis of medical care databases and patients' medical charts. We modified an existing system called MediClass that uses natural language processing (NLP) and knowledge-based methods to classify clinical encounters recorded in electronic medical records (EMRs). We developed the knowledge necessary for MediClass to detect possible vaccine reactions in the outpatient, ED, and telephone encounters recorded in the EMR of a large HMO. We first trained the system using a manually coded gold standard training set, and achieved high sensitivity and specificity. We then ran a large set of post-immunization encounter records through MediClass to see if our method would generalize. Compared to methods that use administrative and clinical codes assigned to the EMR by clinicians, the system significantly improves the positive predictive value for detecting possible vaccine reactions.

40. Raebel MA, Ellis JL, Andrade SE. Evaluation of gestational age and admission date assumptions used to determine prenatal drug exposure from administrative data. Pharmacoepidemiol Drug Saf. 2005 Dec;14(12):829-36. PMID: 15800957.

Objective: *Our aim was to evaluate the 270-day gestational age and delivery date assumptions used in an administrative dataset study assessing prenatal drug exposure compared to information contained in a birth registry.*
Study Design and Setting: *Kaiser Permanente Colorado (KPCO), a member of the Health Maintenance Organization (HMO) Research Network Center for Education and Research in Therapeutics (CERTs), previously participated in a CERTs study that used claims data to assess prenatal drug exposure. In the current study, gestational age and deliveries information from the CERTs study dataset, the Prescribing Safely during Pregnancy Dataset (PSDPD), was compared to information in the KPCO Birth Registry. Sensitivity and positive predictive value (PPV) of the claims data for deliveries were assessed. The effect of gestational age and delivery date assumptions on classification of prenatal drug exposure was evaluated.*
Results: *The mean gestational age in the Birth Registry was 273 (median = 275) days. Sensitivity of claims data at identifying deliveries was 97.6%, PPV was 98.2%. Of deliveries identified in only one dataset, 45% were related to the gestational age assumption and 36% were due to claims data issues. The effect on estimates of prevalence of prescribing during pregnancy was an absolute change of 1% or less for all drug exposure categories. For Category X, drug exposures during the first trimester, the relative change in prescribing prevalence was 13.7% (p = 0.014).*
Conclusion: *Administrative databases can be useful for assessing prenatal drug exposure, but gestational age assumptions can result in a small proportion of misclassification.*

41. Herrinton LJ, Liu L, Lafata JE, Allison JE, Andrade SE, Korner EJ, Chan KA, Platt R, Hiatt D, O'Connor S. Estimation of the period prevalence of inflammatory bowel disease among nine health plans using computerized diagnoses and outpatient pharmacy dispensings. Inflamm Bowel Dis. 2007 Apr;13(4):451-61. PMID: 17219403.

Background: *There are few contemporary estimates of prevalence rates for inflammatory bowel disease (IBD) in diverse North American communities. Methods: We estimated the period prevalence of IBD for January 1, 1999, through*

June 30, 2001, among 1.8 million randomly sampled members of nine integrated healthcare organizations in the US using computerized diagnoses and outpatient pharmaceutical dispensing. We also assessed the positive predictive value (PPV) and sensitivities of 1) the case-finding algorithm, and 2) the 30-month sampling period using medical chart review and linkage to a 78-month dataset, respectively.
Results: *The PPV of the case-finding algorithm was 81% (95% confidence interval [CI], 78-87) and 84% (95% CI, 79-89) in two different organizations. In both, the sensitivity of the optimal algorithm, compared with the most inclusive, exceeded 90%. The sensitivity of the 30-month sampling period compared with 78 months was 61% (95% CI, 57-64) in one organization. Applying a slightly more sensitive case-finding algorithm, the average period prevalence of IBD across the nine organizations, standardized to the age- and gender-distribution of the US population, 2000 census, was 388 cases (95% CI, 378-397) per 100,000 persons (range 209-784 per 100,000; average follow-up 26 months). The prevalence of Crohn's disease, ulcerative colitis, and unspecified IBD was 129, 191, and 69 per 100,000, respectively.*
Conclusions: *The observed average prevalence was similar to prevalence proportions reported for other North American populations (369-408 per 100,000). Additional research is needed to understand differences in the occurrence of IBD among diverse populations as well as practice variation in diagnosis and treatment of IBD.*

42. Mullooly JP, Donahue JG, DeStefano F, Baggs J, Eriksen E. Predictive value of ICD-9-CM codes used in vaccine safety research. *Methods Inf Med* 2008;47:328-335. PMID: 18690366.

   Objectives: *To assess how well selected ICD-9-CM diagnosis codes predict adverse events; to model bias and power loss when vaccine safety analyses rely on unverified codes.*
   Methods: *We extracted chart verification data for ICD-9-CM diagnosis codes from six Vaccine Safety Datalink (VSD) publications and modeled biases and power losses using positive predictive value (PPV) estimates and ranges of code sensitivity.*
   Results: *Positive predictive values were high for type 1 diabetes (80%) in children, relative to WHO criteria, and intussusception (81%) in young children, relative to a standard published case definition. PPVs were moderate (65%) for inpatient and emergency department childhood seizures and low (21%) for outpatient childhood seizures, both relative to physician investigator judgment. Codes for incident central nervous system demyelinating disease in adults had high PPV for inpatient codes (80%) and low PPV for outpatient codes (42%) relative to physicians' diagnoses. Modeled biases were modest, but large increases in frequencies of adverse events are required to achieve adequate power if unverified ICD-9-CM codes are used, especially when vaccine associations are weak.*
   Conclusions: *ICD-9-CM codes for type 1 diabetes in children, intussusception in young children, childhood seizures in inpatient and emergency care settings, and inpatient demyelinating disease in adults were sufficiently predictive for vaccine safety analyses to rely on unverified diagnosis codes. Adverse event misclassification should be accounted for in statistical power calculations.*

43. Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, et al. Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. Med Care.2009;47:1091–1097. PMID: 19648826.

Background: Chemotherapy data are important to almost any study on cancer prognosis and outcomes. However, chemotherapy data obtained from tumor registries may be incomplete, and abstracting chemotherapy directly from medical records can be expensive and time consuming.
Methods: We evaluated the accuracy of using automated clinical data to capture chemotherapy administrations in a cohort of 757 ovarian cancer patients enrolled in 7 health plans in the HMO Cancer Research Network. We calculated sensitivity and specificity with 95% confidence intervals of chemotherapy administrations extracted from 3 automated clinical data sources (Health Care Procedure Coding System, National Drug Codes, and International Classification of Diseases) compared with tumor registry data and medical chart data.
Results: Sensitivity of all 3 data sources varied across health plans from 79.4% to 95.2% when compared with tumor registries, and 75.0% to 100.0% when compared with medical charts. The sensitivities using a combination of 3 data sources were 88.6% (95% confidence intervals: 85.7-91.1) compared with tumor registries and 89.5% (78.5-96.0) compared with medical records; specificities were 91.5% (86.4-95.2) and 90.0% (55.5-99.7), respectively. There was no difference in accuracy between women aged < 65 and > or = 65 years. Using one set of codes alone (eg, Health Care Procedure Coding System alone) was insufficient for capturing chemotherapy data at most health plans.
Conclusions: While automated data systems are not without limitations, clinical codes used in combination are useful in capturing chemotherapy more comprehensively than tumor registry and without the need for costly medical record abstraction.

44. Shui IM, Shi P, Dutta-Linn MM, Weintraub ES, Hambidge SJ, Nordin JD, Lieu TA. Predictive value of seizure ICD-9 codes for vaccine safety research. *Vaccine* 2009;27:5307-5312. PMID: 19616500.

Post-licensure vaccine safety studies often monitor for seizures using automated screening of ICD-9 codes. This study assessed the positive predictive value (PPV) of ICD-9 codes used to identify seizure visits in children aged 6 weeks to 23 months who were enrolled in seven managed care organizations during January 2000 to December 2005. ICD-9 codes were used to identify visits for seizures in the 0-30-day period following receipt of a pneumococcal vaccine. Visits were stratified by setting of diagnosis (emergency department (ED), outpatient, and inpatient). Review of medical records confirmed whether the visit represented a true acute seizure event. 3233 visits for seizures were identified; 1024 were randomly selected for medical record review and 859 (84%) had records available. The PPV of ICD-9 codes was highest in the ED setting (97%), followed by the inpatient setting (64%). In the outpatient setting, computerized codes for seizures had very low PPV: 16% on days 1-30 following vaccination and 2% for visits on the same day of vaccination. An estimated 77% of true seizures identified were from the ED or inpatient settings. In conclusion, when using ICD-9 codes to identify seizure outcomes, restricting to the ED and inpatient settings of diagnosis may result in less biased preliminary analyses and more efficient vaccine safety studies.

45. Greene SK, Shi P, Dutta-Linn MM, Shoup JA, Hinrichsen VL, Ray P, Nordin JD, Kuckler L, Weintraub ES., et al. Accuracy of data on influenza vaccination status at four Vaccine Safety Datalink sites. *Am J Prev Med* 2009;37:552-555. PMID: 1994924.

Background: Studies of influenza vaccination using electronic medical records rely on accurate classification of vaccination status. Vaccinations not entered into electronic records would be unavailable for study.

Purpose: This study evaluated the sensitivity and negative predictive value (NPV) of electronic records for influenza vaccination and factors associated with failure to capture vaccinations.

Methods: In four diverse medical care organizations in the Vaccine Safety Datalink, those aged 50-79 years with no influenza vaccination record during the 2007-2008 season were surveyed by telephone, and electronic records were analyzed in 2008. The sensitivity and NPV of electronic records were estimated, using survey responses as the gold standard. Logistic regression models determined associations between 1-NPV and demographic factors, risk of influenza complications, and healthcare utilization levels.

Results: Data were obtained for 933 survey participants and 1,085,916 medical care organization members. Sites varied significantly in the sensitivity (51%, 68%, 79%, 89%) and NPV (46%, 62%, 66%, 87%) of electronic records. In multivariate analysis, the rate of failure to capture vaccinations was significantly higher for those aged 65-79 years than for those aged 50-64 years at three sites. Of vaccinations not captured by electronic records, 58% were reportedly administered in nontraditional settings, usually workplaces; the rest were given within the sites.

Conclusions: Influenza vaccination studies relying on electronic records may misclassify substantial proportions of vaccinated individuals as unvaccinated, producing biased estimates of vaccine effectiveness. Sites with limited sensitivity to capture vaccinations administered within their organization should seek possible remedies. More complete capture of vaccinations administered to older patients and in nontraditional settings would further reduce misclassification.

46. Arterburn DE, Alexander GL, Calvi J, Coleman LA, Gillman MW, Novotny R, Quinn VP, Rukstalis M, Stevens VJ, Taveras EM, Sherwood NE. Body mass index measurement and obesity prevalence in ten U.S. health plans. Clin Med Res. 2010 Dec;8(3-4):126-30. PMID: 20682758.

Objective: The objective of this study was to examine the frequency of body mass index (BMI) measurement before the implementation of two new Healthcare Effectiveness Data and Information Set (HEDIS) performance measures for obesity that require U.S. health plans to annually report the frequency of BMI and BMI percentile measurement among all adults and children who had at least one outpatient visit during the past two years.

Design: Cross-sectional study.

Setting: A consortium of ten U.S. health plans and care delivery systems from the Health Maintenance Organization Research Network, which together provide care to more than 6.5 million adults and children.

Participants: Children and adults, age 2 years and older, who were continuously enrolled in one of ten U.S. health plans for at least one full year from 2005 to 2006

Methods: We extracted available anthropometric data for 3.7 million adults and 1.2 million children with at least one visit captured from ten electronic medical record databases from 2005 to 2006.

Results: We found that the availability of BMI measurements for adults ranged widely across health plans from 28% to 88%, and availability of BMI percentiles for children ranged from 21% to 81%. Among adults and children with BMI measures in

these ten health plans, the overall prevalence of overweight and obesity were very similar to those reported in the 2005 to 2006 U.S. national surveys that used measured heights and weights.

Conclusion: *The newly approved HEDIS performance measures likely represent an important step in addressing the quality of obesity care in the United States. The current study demonstrates that these HEDIS measures are achievable, especially among health plans that have implemented electronic medical records. Future research should assess the relationship between BMI assessment, provider counseling and treatment practices, and long-term changes in obesity rates among different population groups.*

47. Smith N, Iyer RL, Langer-Gould A, Getahun DT, Strickland D, Jacobsen SJ, Chen W, Derose SF, Koebnick C. Health plan administrative records versus birth certificate records: quality of race and ethnicity information in children. BMC Health Serv Res. 2010 Nov 23;10:316. PMID: 21092309.

Background: *To understand racial and ethnic disparities in health care utilization and their potential underlying causes, valid information on race and ethnicity is necessary. However, the validity of pediatric race and ethnicity information in administrative records from large integrated health care systems using electronic medical records is largely unknown.*

Methods: *Information on race and ethnicity of 325,810 children born between 1998-2008 was extracted from health plan administrative records and compared to birth certificate records. Positive predictive values (PPV) were calculated for correct classification of race and ethnicity in administrative records compared to birth certificate records.*

Results: *Misclassification of ethnicity and race in administrative records occurred in 23.1% and 33.6% children, respectively; the majority due to missing ethnicity (48.3%) and race (40.9%) information. Misclassification was most common in children of minority groups. PPV for White, Black, Asian/Pacific Islander, American Indian/Alaskan Native, multiple and other was 89.3%, 86.6%, 73.8%, 18.2%, 51.8% and 1.2%, respectively. PPV for Hispanic ethnicity was 95.6%. Racial and ethnic information improved with increasing number of medical visits. Subgroup analyses comparing racial classification between non-Hispanics and Hispanics showed White, Black and Asian race was more accurate among non-Hispanics than Hispanics.*

Conclusions: *In children, race and ethnicity information from administrative records has significant limitations in accurately identifying small minority groups. These results suggest that the quality of racial information obtained from administrative records may benefit from additional supplementation by birth certificate data.*

48. Andrade SE, Moore Simas TA, Boudreau D, Raebel MA, Toh S, Syat B, Dashevsky I, Platt R.   Validation of algorithms to ascertain clinical conditions and medical procedures used during pregnancy. Pharmacoepidemiol Drug Saf. 2011 Nov;20(11):1168-76. PMID: 22020902.

Purpose: *To evaluate the validity of health plan administrative and claims data to identify pre-gestational and gestational diabetes, obesity, and ultrasounds among pregnant women.*

Methods: *A retrospective study was conducted using the administrative and claims data of three health plans participating in the HMO Research Network. Diagnoses, drug dispensings, and procedure codes were used to identify diabetes, obesity, and*

ultrasounds among women who were pregnant between January 2006 and December 2008. A random sample of medical charts (n=222) were abstracted. Positive predictive values (PPVs) were calculated. Sensitivity also was calculated for obesity among women for whom body mass index data were available in electronic medical records at two sites.
***Results:** Overall, 190 of 222 cases of diabetes (86%) were confirmed (82% for gestational diabetes and 74% for pre-gestational diabetes). The PPV for codes to identify ultrasounds was 80%. Whereas the PPV for obesity-related diagnosis codes was high (93%), and the sensitivity was low (33%).*
***Conclusions:** Health plan administrative and claims data can be used to accurately identify pre-gestational and gestational diabetes and ultrasounds. Obesity is not consistently coded.*

49. Scholes D, Yu O, Raebel MA, Trabert B, Holt VL. Improving automated case finding for ectopic pregnancy using a classification algorithm. Hum Reprod. 2011 Nov;26(11):3163-8. PMID: 21911435.

    ***Background:** Research and surveillance work addressing ectopic pregnancy often rely on diagnosis and procedure codes available from automated data sources. However, the use of these codes may result in misclassification of cases. Our aims were to evaluate the accuracy of standard ectopic pregnancy codes; and, through the use of additional automated data, to develop and validate a classification algorithm that could potentially improve the accuracy of ectopic pregnancy case identification.*
    ***Methods:** Using automated databases from two US managed-care plans, Group Health Cooperative (GH) and Kaiser Permanente Colorado (KPCO), we sampled women aged 15-44 with an ectopic pregnancy diagnosis or procedure code from 2001 to 2007 and verified their true case status through medical record review. We calculated positive predictive values (PPV) for code-selected cases compared with true cases at both sites. Using additional variables from the automated databases and classification and regression tree (CART) analysis, we developed a case-finding algorithm at GH (n = 280), which was validated at KPCO (n = 500).*
    ***Results:** Compared with true cases, the PPV of code-selected cases was 68 and 81% at GH and KPCO, respectively. The case-finding algorithm identified three predictors: ≥ 2 visits with an ectopic pregnancy code within 180 days; International Classification of Diseases, 9th Revision, Clinical Modification codes for tubal pregnancy; and methotrexate treatment. Relative to true cases, performance measures for the development and validation sets, respectively, were: 93 and 95% sensitivity; 81 and 81% specificity; 91 and 96% PPV; 84 and 79% negative predictive value. Misclassification proportions were 32% in the development set and 19% in the validation set when using standard codes; they were 11 and 8%, respectively, when using the algorithm.*
    ***Conclusions:** The ectopic pregnancy algorithm improved case-finding accuracy over use of standard codes alone and generalized well to a second site. When using administrative data to select potential ectopic pregnancy cases, additional widely available automated health plan data offer the potential to improve case identification.*

50. Aiello Bowles EJ, Feigelson HS, Barney T, Broecker K, Sterrett A, Bischoff K, Engel J, Gundersen G, Sheehey-Jones J., et al. Improving quality of breast cancer surgery through development of a national breast cancer surgical outcomes (BRCASO) research database. *BMC Cancer* 2012;12:136. PMID: 22472011.

*BACKGROUND: Common measures of surgical quality are 30-day morbidity and mortality, which poorly describe breast cancer surgical quality with extremely low morbidity and mortality rates. Several national quality programs have collected additional surgical quality measures; however, program participation is voluntary and results may not be generalizable to all surgeons. We developed the Breast Cancer Surgical Outcomes (BRCASO) database to capture meaningful breast cancer surgical quality measures among a non-voluntary sample, and study variation in these measures across providers, facilities, and health plans. This paper describes our study protocol, data collection methods, and summarizes the strengths and limitations of these data. **Methods:** We included 4524 women >/=18 years diagnosed with breast cancer between 2003-2008. All women with initial breast cancer surgery performed by a surgeon employed at the University of Vermont or three Cancer Research Network (CRN) health plans were eligible for inclusion. From the CRN institutions, we collected electronic administrative data including tumor registry information, Current Procedure Terminology codes for breast cancer surgeries, surgeons, surgical facilities, and patient demographics. We supplemented electronic data with medical record abstraction to collect additional pathology and surgery detail. All data were manually abstracted at the University of Vermont. **Results:** The CRN institutions pre-filled 30% (22 out of 72) of elements using electronic data. The remaining elements, including detailed pathology margin status and breast and lymph node surgeries, required chart abstraction. The mean age was 61 years (range 20-98 years); 70% of women were diagnosed with invasive ductal carcinoma, 20% with ductal carcinoma in situ, and 10% with invasive lobular carcinoma. **Conclusions:** The BRCASO database is one of the largest, multi-site research resources of meaningful breast cancer surgical quality data in the United States. Assembling data from electronic administrative databases and manual chart review balanced efficiency with high-quality, unbiased data collection. Using the BRCASO database, we will evaluate surgical quality measures including mastectomy rates, positive margin rates, and partial mastectomy re-excision rates among a diverse, non-voluntary population of patients, providers, and facilities.*

51. Ritzwoller DP, Carroll N, Delate T, O'Keeffe-Rossetti M, Fishman PA, Loggers ET, Aiello Bowles EJ, Elston Lafata J, Hornbrook MC. Validation of Electronic Data on Chemotherapy and Hormone Therapy Use in HMOs. Med Care. 2012 Apr 23. [Epub ahead of print]. PMID: 22531648.

Background: Most data regarding medical care for cancer patients in the United States comes from Surveillance, Epidemiology and End Results-linked Medicare analyses of individuals aged 65 years or older and typically excludes Medicare Advantage enrollees.
Objectives: To assess the accuracy of chemotherapy and hormone therapy treatment data available through the Cancer Research Network's Virtual Data Warehouse (VDW).
Research Design: Retrospective, longitudinal cohort study. Medical record-abstracted, tumor registry-indicated treatments (gold standard) were compared with VDW-indicated treatments derived from health maintenance organization pharmacy, electronic medical record, and claim-based data systems.
Subjects: Enrollees aged 18 years and older diagnosed with incident breast, colorectal, lung, or prostate cancer from 2000 through 2007.

Measures: Sensitivity, specificity, and positive predictive value were computed at 6 and 12 months after cancer diagnosis.
Results: Approximately 45% of all cancer cases (total N=23,800) were aged 64 years or younger. Overall chemotherapy sensitivity/specificities across the 3 health plans for incident breast, colorectal, lung, and prostate cancer cases were 95%/90%, 95%/93%, 93%/93%, and 85%/77%, respectively. With the exception of prostate cancer cases, overall positive predictive value ranged from 86% to 89%. Small variations in chemotherapy data accuracy existed due to cancer site and data source, whereas greater variation existed in hormone therapy capture across sites.
Conclusions: Strong concordance exists between gold standard tumor registry measures of chemotherapy receipt and Cancer Research Network VDW data. Health maintenance organization VDW data can be used for a variety of studies addressing patterns of cancer care and comparative effectiveness research that previously could only be conducted among elderly Surveillance, Epidemiology and End Results-Medicare populations.

52. Kahn MG, Raebel, MA, Glanz JM, Riedlinger K, Steiner, JF. A Pragmatic Framework for Single-site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research. Med Care 2012 Jul; 50 Suppl: S21–S29. PMID: 22692254.

Introduction: Answers to clinical and public health research questions increasingly require aggregated data from multiple sites. Data from electronic health records and other clinical sources are useful for such studies, but require stringent quality assessment. Data quality assessment is particularly important in multisite studies to distinguish true variations in care from data quality problems.
Methods: We propose a "fit-for-use" conceptual model for data quality assessment and a process model for planning and conducting single-site and multisite data quality assessments. These approaches are illustrated using examples from prior multisite studies.
Approach: Critical components of multisite data quality assessment include: thoughtful prioritization of variables and data quality dimensions for assessment; development and use of standardized approaches to data quality assessment that can improve data utility over time; iterative cycles of assessment within and between sites; targeting assessment toward data domains known to be vulnerable to quality problems; and detailed documentation of the rationale and outcomes of data quality assessments to inform data users. The assessment process requires constant communication between site-level data providers, data coordinating centers, and principal investigators.
Discussion: A conceptually based and systematically executed approach to data quality assessment is essential to achieve the potential of the electronic revolution in health care. High-quality data allow "learning health care organizations" to analyze and act on their own information, to compare their outcomes to peers, and to address critical scientific questions from the population perspective.

53. Allen LA, Yood MU, Wagner EH, Aiello Bowles EJ, Pardee R, Wellman R, Habel L, Nekhlyudov L, Davis RL, Onitilo AA, Magid DJ; for the Pharmacovigilance Research Group. Performance of Claims-based Algorithms for Identifying Heart Failure and Cardiomyopathy among Patients Diagnosed with Breast Cancer. Med Care. 2012 May 25. [Epub ahead of print]. PMID: 22643199.

Background: *Cardiotoxicity is a known complication of certain breast cancer therapies, but rates come from clinical trials with design features that limit external validity. The ability to accurately identify cardiotoxicity from administrative data would enhance safety information.*

Objective: *To characterize the performance of clinical coding algorithms for identification of cardiac dysfunction in a cancer population.*

Research Design: *We sampled 400 charts among 6460 women diagnosed with incident breast cancer, tumor size ≥2 cm or node positivity, treated within 8 US health care systems between 1999 and 2007. We abstracted medical records for clinical diagnoses of heart failure (HF) and cardiomyopathy (CM) or evidence of reduced left ventricular ejection fraction. We then assessed the performance of 3 different International Classification of Diseases, 9th Edition (ICD-9)-based algorithms.*

Results: *The HF/CM coding algorithm designed a priori to balance performance characteristics provided a sensitivity of 62% (95% confidence interval, 40%-80%), specificity of 99% (range, 97% to 99%), positive predictive value (PPV) of 69% (range, 45% to 85%), and negative predictive value (NPV) of 98% (range, 96% to 99%). When applied only to incident HF/CM (ICD-9 codes and gold standard diagnosis both occurring after breast cancer diagnosis) in patients exposed to anthracycline and/or trastuzumab therapy, the PPV was 42% (range, 14% to 76%).*

Conclusions: *Claims-based algorithms have moderate sensitivity and high specificity for identifying HF/CM among patients with invasive breast cancer. As the prevalence of HF/CM among the breast cancer population is low, ICD-9 codes have high NPV but only moderate PPV. These findings suggest a significant degree of misclassification due to HF/CM overcoding versus incomplete clinical documentation of HF/CM in the medical record.*

54. Delate T, Bowles EJ, Pardee R, Wellman RD, Habel LA, Yood MU, Nekhlyudov L, Goddard KA, Davis RL, McCarty CA, Onitilo AA, Feigelson HS, Freml J, Wagner E. Validity of eight integrated healthcare delivery organizations' administrative clinical data to capture breast cancer chemotherapy exposure. Cancer Epidemiol Biomarkers Prev. 2012 Apr;21(4):673-80. PMID: 22337532.

Background: *Cancer Research Network (CRN) sites use administrative data to populate their Virtual Data Warehouse (VDW). However, information on VDW chemotherapy data validity is limited. The purpose of this study was to assess the validity of VDW chemotherapy data.*

Methods: *This was a retrospective cohort study of women ≥18 years with incident, invasive breast cancer diagnosed between January 1999 and December 2007. Pharmacy and procedure chemotherapy data were extracted from each site's VDW. Random samples of 50 patients stratified on trastuzumab, anthracyclines, and no chemotherapy exposure was selected from each site for detailed chart abstraction. Weighted sensitivities and specificities of VDW compared with abstracted data were calculated. Cumulative doses calculated from VDW data were compared with doses obtained from the medical chart review.*

Results: *The cohort included 13,497 patients with 6,456 (48%) chart review eligible. Patients in the sample (N = 400) had a mean age of 65 years. Trastuzumab, anthracycline, and other chemotherapy weighted sensitivities were 95%, 97%, and 100%, respectively; specificities were 99%, 99%, and 93%, respectively; positive predictive values were 96%, 99%, and 55%, respectively; and negative predictive values were 99%, 96%, and 100%. Trastuzumab and anthracyclines VDW mean*

doses were 873 and 386 mg, respectively, whereas abstracted mean doses were 1,734 and 369 mgs, respectively (R(2) = 0.14, P < 0.01 and R(2) = 0.05, P = 0.03, respectively).
Conclusions: Sensitivities and specificities for CRN chemotherapy VDW data were high and dosages were correlated with chart information.
IMPACT: The findings support the use of CRN data in evaluating chemotherapy exposures and related outcomes.

V.   Human Subjects' Privacy and Safety

55. Greene SM, Geiger AM. A review finds multicenter studies face substantial challenges but strategies exist to achieve Institutional Review Board approval. J Clin Epidemiol. 2006;59:784-790. PMID: 16828670.

    Background and Objectives: To characterize the impact of multiple Institutional Review Board (IRB) reviews on multicenter observational research studies, and identify strategies for overcoming the identified challenges.
    Materials And Methods: Using PubMed, we identified empirical studies, anecdotal reports, and opinion pieces addressing the process of obtaining initial IRB approval for multicenter clinical trials and observational studies. We also reviewed relevant information from federal and other national sources.
    Results: A total of 40 peer-reviewed articles were synthesized, plus six reports from commissions or other key sources. These sources identified numerous challenges that researchers may encounter when multicenter studies undergo review by multiple IRBs, such as added time for initial review and approval and different requirements across IRBs. Strategies to alleviate these challenges include planning to accommodate multiple reviews and determining upfront whether certain variations to study protocols are tolerable across sites. Many researchers and commissions have proposed comprehensive reforms, such as centralized review for multicenter projects.
    Conclusion: Policy-makers, researchers, and IRBs should convene to specifically discuss optimal approaches for multicenter review. However, until structural changes are implemented, observational researchers should develop and implement strategic plans for obtaining IRB approval in multicenter studies, including adopting models successfully employed by clinical trials.

56. Greene SM, Geiger AM, Harris EL, Altschuler A, Nekhlyudov L, Barton MB, Rolnick SJ, Elmore JG, Fletcher S. Impact of IRB requirements on a multicenter survey of prophylactic mastectomy outcomes. *Ann Epidemiol* 2006;16:275-278. PMID: 16005245.

    Objective: This study assesses the variability in requirements among six institutional review boards (IRBs) and the resulting protocol variations for a multicenter mailed survey.
    Study Design and Setting: We utilized a cross-sectional mailed survey to gather information on long-term psychosocial outcomes of prophylactic mastectomy among women at six health maintenance organizations, all of which are part of the Cancer Research Network. In the context of this collaborative study, we characterized the impact of the different sites' IRB review processes on the study protocol and participation.

Results: IRB review resulted in site differences in physician consent prior to participant contact, invitation letter content and signatories, and incentive type. The review process required two to eight modifications beyond the initial application and resulted in unanticipated delays and costs.
Conclusion: Site-to-site variability in IRB requirements may adversely impact scientific rigor and delay implementation of collaborative studies, especially when not considered in project planning. IRB review is an essential aspect of research but one that can present substantial challenges for multicenter studies.

57. Greene SM, Braff J, Nelson AF, Reid RJ. The Process Is the Product: A New Model for Multisite IRB Review of Data-Only Studies. IRB: Ethics & Human Research, Vol 32(3), May-Jun 2010, 1-6. **No abstract.** PMID: 20590050.

58. Greene,S.M.; Braff,J.; Nelson,A.F.; Reid,R.J. HMO group uses alternate IRB model to review low-risk studies. IRB Advisor. 11(3):30-31, March 2011. **No abstract.**

59. Marsolo, K. Approaches to facilitate Institutional Review Board Approval of Multicenter Research Studies. Med Care 2012 Jul; 50 Suppl: S77–S81. PMID: 22692264.

Background and Objectives: Gaining Institutional Review Board (IRB) approval for a multicenter research study can be a lengthy and time-consuming process. It can increase the complexity of consent forms, decreasing patient understanding and lowering recruitment numbers. It also leads to increased costs through the duplication of effort. This paper examines some of the strategies used to streamline the IRB review process for multicenter studies and provides examples used by 2 existing multicenter comparative effectiveness research networks.
Methods: A literature search was conducted to identify sources that described the challenges and potential strategies to facilitate multicenter IRB approval. The most promising avenues were identified and included in this review. Phone interviews were conducted with the Principal Investigators and Project Managers of 2 successful multicenter research networks to learn their "keys to success" and their lessons learned.
Results: Three strategies were identified that held the most promise: working with IRBs before submission, the use of central and/or federated IRBs, and the establishment of an umbrella protocol. Each of these strategies was used to some degree by the case study projects.
Conclusions: Although the approaches documented here can help streamline the IRB approval process, they are not a "silver bullet." Because some of these approaches are still relatively new, empirical data are sparse. However, it is believed that they will significantly reduce the administrative burden of the project as a whole and lead to a decrease in the overall time to protocol approval.

VI. HMORN Research Methods

60. Field TS, Cadoret CA, Brown ML, Ford M, Greene SM, Hill D, Hornbrook MC, MeenanRT, White MJ, Zapka JM. Surveying physicians: do components of the "Total Design Approach" to optimizing survey response rates apply to physicians? Med Care. 2002 Jul;40(7):596-605. PMID: 12142775.

Background: Surveys serve essential roles in clinical epidemiology and health services research. However, physician surveys frequently encounter problems achieving adequate response rates. Research on enhancing response rates to surveys of the general public has led to the development of Dillman's "Total Design Approach" to the design and conduct of surveys. The impact of this approach on response rates among physicians is uncertain.
Objective: To determine the extent to which the components of the total design approach have been found to be effective in physician surveys.
Design: A systematic review.
Results: The effectiveness of prepaid financial incentives, special contacts, and personalization to enhance response rates in surveys of physicians have been confirmed by the existing research. There is suggestive evidence supporting the use of first class stamps on return envelopes and multiple contacts. The optimum amount for incentives and the number of contacts necessary have not been established. Details of questionnaire design and their impact on response rates have received almost no attention from researchers. Few studies have assessed the usefulness of combinations of components of the total design approach.
Conclusions: Despite the number of surveys conducted among physicians, their cost, the level of interest in their findings, and in spite of inadequate response rates, there have been few randomized trials conducted on important aspects of enhancing response in this population. Until this gap has been filled, researchers conducting surveys of physicians should consider including all components of the total design approach whenever feasible.

61. Puleo E, Zapka J, White MJ, Mouchawar J, Somkin C, Taplin S. Caffeine, cajoling, and other strategies to maximize clinician survey response rates. *Eval Health Prof* 2002;25:169-184. PMID: 12026751.

An ongoing objective in health services research is to increase response rates to clinician surveys to ensure generalizability of findings. Three HMOs in the Cancer Research Network participated in a primary care clinician survey to better understand organizational characteristics affecting adoption and implementation of breast and cervical cancer screening guidelines. A four-stage data collection strategy was implemented to maximize response. This included careful attention to survey design and layout, extensive piloting, choice of token incentive, use of "local champions," and denominator management. An overall response rate of 91% was attained, ranging from 83 to 100% among the plans (N = 621). Although the response rate after the second stage of data collection met commonly used standards, the authors argue for the four-stage method due to the possibility of differences when comparing early and late responders. This is important when multiple plans with differing structure and internal characteristics are surveyed.

62. Reisch LM, Fosse JS, Beverly K, Yu O, Barlow WE, Harris EL, et al. Training, quality assurance, and assessment of medical record abstraction in a multisite study. Am J Epidemiol 2003; 157: 546–51. PMID: 12631545.

Clinical studies using medical record review should include careful training and quality assurance methods to enhance the reliability and validity of data obtained from the records. Because of time and budget constraints, comprehensive assessments of data quality and reliability, including masking of medical record abstractors, are not always possible. This paper describes the abstractor training and quality control methods and results of a masked medical record review study. The medical record review study was carried out within a larger multisite study of the effectiveness of screening mammography in preventing breast cancer mortality with an observation period within 1983 and 1993, with mortality follow-up through 1998. An eight-step program was developed to train medical record abstractors and monitor the quality of their work. A key follow-up component to the training protocol was a 5% reabstraction of medical records (n = 160), masked and reviewed by a second abstractor. High agreement was found between initial (unmasked) abstractors and masked abstractors for all key exposure variables (kappa ranged from 0.76 to 0.91), with no evidence of biased directionality by unmasked reviewers. Rigorous ongoing training programs for medical record abstractors provide assurance of good quality control in large multisite studies. Additionally, a masking study with a subsample of subjects may be a feasible and cost-effective alternative to the time- and cost-intensive methodological approach of masking all medical records.

63. Field TS, Cernieux J, Buist D, Geiger A, Lamerato L, Hart G, Bachman D, Krajenta R, Greene S, Hornbrook MC, Ansell G, Herrinton L, Reed G. Retention of enrollees following a cancer diagnosis within health maintenance organizations in the Cancer Research Network. J Natl Cancer Inst. 2004 Jan 21;96(2):148-52. PMID: 14734705.

Population laboratories with complete clinical information on episodes of care are needed to support research on the quality of care delivered to cancer patients. Data resources within the Cancer Research Network (CRN) may overcome many of the limitations of existing cancer databases, but their potential clinical value depends on the stability of the enrolled population. To assess this issue, we studied the retention rates among survivors of the 132 580 patients diagnosed with cancer from January 1, 1993, through December 31, 1998, who were enrolled at five health maintenance organization sites participating in the CRN. Enrollees were followed from cancer diagnosis through death, disenrollment, or the end of follow-up (i.e., December 31, 1999). The retention rate among survivors for all cancers combined at 1 and 5 years after cancer diagnosis was 96.0% (95% confidence interval [CI] = 95.9% to 96.1%) and 83.9% (95% CI = 83.4% to 84.3%), respectively. The proportion of enrollees diagnosed with cancer who remained enrolled and available for evaluation suggests that the CRN is well-suited for studies of the quality of care for cancer patients, survivorship, and long-term outcomes.

64. Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay SK, Platt R, Black S, Shinefield H, Chen RT. Active Surveillance of Vaccine Safety. A System to Detect Early Signs of Adverse Events. Epidemiology 2005;May;16(3):336-341. PMID: 15824549.

Background: There currently are no population-based systems in the United States to rapidly detect adverse events after newly introduced vaccines. To evaluate the feasibility of developing such systems, we used 5 years of data from 4 health maintenance organizations within the Centers for Disease Control and Prevention (CDC) Vaccine Safety Datalink.
Methods: Within every year, each week's vaccinated children were followed for 4 weeks, and rates of adverse events were compared with rates among children of similar ages before the introduction of the new vaccine. We assessed risks for intussusception after rotavirus vaccination and risks for fever, seizures, and other neurologic adverse events after the change from whole cell diphtheria-tetanus pertussis (DTPw) to acellular DTP vaccine (DTPa). We used sequential probability ratio testing, adjusted for age, sex, calendar time, season, and HMO, and with a stopping value based on the probability of an adverse event under the null hypothesis and under a preset alternative hypothesis.
Results: We detected an increase in intussusception after 2589 vaccine doses of rotavirus vaccine, about the same time initial reports of intussusception were made to the Vaccine Adverse Events Reporting System. Decreases in risk for fever, seizures, and other abnormal neurologic events became detectable within 12 weeks, 42 weeks, and 18 months, respectively, after the change from DTPw to DTPa.

65. Mazor KM, Sabin JE, Boudreau D, Goodman MJ, Gurwitz JH, Herrinton LJ, Raebel MA, Roblin D, Smith DH, Meterko V, Platt R. Cluster randomized trials: opportunities and barriers identified by leaders of eight health plans. Med Care. 2007 Oct;45(10 Supl 2):S29-37. PMID: 17909379

Background: Cluster randomized trials (CRTs) offer unique advantages over standard randomized controlled clinical trials (RCTs) and observational methodologies, and may provide a cost-efficient alternative for answering questions about the best treatments for common conditions.
Objectives: To describe health plan leaders' views on CRTs, identify barriers to conducting CRTs, and solicit recommendations for increasing the acceptability of CRTs.
Research Design: Qualitative in-depth telephone interviews with leaders from 8 health plans.
Subjects: Thirty-four health plan leaders (medical directors, pharmacy directors, Institutional Review Board leaders, ethics leaders, compliance leaders, and others).
Measures: Qualitative analysis of interview transcripts to identify barriers, factors influencing leaders' views, ethical issues, aspects of CRTs that appeal to leaders, and recommendations for increasing acceptability of CRTs.
Results: Multiple barriers were identified, including financial costs, concerns about stakeholders' perceptions of CRTs, impact on physicians' prescribing habits, and formulary changes. Most leaders recognized the potential value of studying the comparative effectiveness of therapeutics, and many stressed the need for head-to-head trials. Leaders' views would be influenced by variations in study design and implementation. Recommendations for increasing acceptability of CRTs included

ensuring that the fiscal impact of a CRT be budget neutral, and that researchers educate stakeholders and decision-makers about CRTs.
Conclusions: *Overall, health plan leaders recognized the need for studies of the comparative effectiveness of therapeutics under real world conditions, and many expressed support for CRTs. However, researchers seeking to conduct CRTs in health plans are likely to face numerous barriers, and preparatory work will be essential.*

66. Thwin SS, Clough-Gorr KM, McCarty MC, Lash TL, Alford SH, Buist DS, Enger SM, Field TS, Frost F., et al. Automated inter-rater reliability assessment and electronic data collection in a multi-center breast cancer study. *BMC Med Res Methodol* 2007;7:23. PMID: 17577410.

    Background: *The choice between paper data collection methods and electronic data collection (EDC) methods has become a key question for clinical researchers. There remains a need to examine potential benefits, efficiencies, and innovations associated with an EDC system in a multi-center medical record review study.*
    Methods: *A computer-based automated menu-driven system with 658 data fields was developed for a cohort study of women aged 65 years or older, diagnosed with invasive histologically confirmed primary breast cancer (N = 1859), at 6 Cancer Research Network sites. Medical record review with direct data entry into the EDC system was implemented. An inter-rater and intra-rater reliability (IRR) system was developed using a modified version of the EDC.*
    Results: *Automation of EDC accelerated the flow of study information and resulted in an efficient data collection process. Data collection time was reduced by approximately four months compared to the project schedule and funded time available for manuscript preparation increased by 12 months. In addition, an innovative modified version of the EDC permitted an automated evaluation of inter-rater and intra-rater reliability across six data collection sites.*
    Conclusion: *Automated EDC is a powerful tool for research efficiency and innovation, especially when multiple data collection sites are involved.*

67. Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, Yin R, Brown JS, Platt R. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007;45:S89-S95. PMID: 17909389.

    Background: *Rare but serious adverse events associated with vaccines or drugs are often nearly impossible to detect in prelicensure studies and require monitoring after introduction of the agent in large populations. Sequential testing procedures are needed to detect vaccine or drug safety problems as soon as possible after introduction.*
    Objective: *To develop and evaluate a new real-time surveillance system that uses dynamic data files and sequential analysis for early detection of adverse events after the introduction of new vaccines.*
    Research Design: *The Centers for Disease Control and Prevention (CDC)-sponsored Vaccine Safety Datalink Project developed a real-time surveillance system and initiated its use in an ongoing study of a new meningococcal vaccine for adolescents. Dynamic data files from 8 health plans were updated and aggregated for analysis every week. The analysis used maximized sequential probability ratio testing (maxSPRT), a new signal detection method that supports continuous or time-period analysis of data as they are collected.*

Results: Using the new real-time surveillance system, ongoing analyses of meningococcal conjugate vaccine (MCV) safety are being conducted on a weekly basis. Two forms of maxSPRT were implemented: an analysis using concurrent matched controls, and an analysis based on expected counts of the outcomes of interest, which were estimated based on historical data. The analysis highlights both theoretical and operational issues, including how to (1) choose appropriate outcomes and stopping rules, (2) select control groups, and (3) accommodate variation in exposed:unexposed ratios between time periods and study sites.
Conclusions: Real-time surveillance combining dynamic data files, aggregation of data, and sequential analysis methods offers a useful and highly adaptable approach to early detection of adverse events after the introduction of new vaccines.

68. Lash TL, Fox MP, Thwin SS, Geiger AM, Buist DS, Wei F, Field TS, Yood MU, Frost FJ., et al. Using probabilistic corrections to account for abstractor agreement in medical record reviews. *Am J Epidemiol* 2007;165:1454-1461. PMID: 17406006.

The quality of medical record abstracts is often characterized in a reliability substudy. These results usually indicate agreement, but not the extent to which lack of agreement affects associations observed in the complete data. In this study, medical records were reviewed and abstracted for patients diagnosed with stage I or stage II breast cancer between 1990 and 1994 at one of six US Cancer Research Network sites. For a subsample, interrater reliability data were available. The authors calculated conventional hazard ratios and 95% confidence intervals for the association of demographic, tumor, and treatment characteristics with recurrence rate. These conventional estimates of effect were compared with three sets of estimates and 95% simulation intervals that took account of the uncertainty assessed by lack of agreement in the reliability substudy. The rate of recurrence was associated with increasing cancer stage and with treatment modality but not with demographic characteristics. The hazard ratios and simulation intervals that took account of the reliability data showed that the simulation interval grew wider as the sources of uncertainty taken into account grew more complete, but the associations expected a priori remained readily apparent. While many investigators use reliability data only as a metric for data quality, a more thorough approach can also quantitatively depict the uncertainty in the observed associations.

69. Aiello EJ, Buist DS, Wagner EH, Tuzzio L, Greene SM, Lamerato LE, Field TS, Herrinton LJ, Haque R, Hart G, Bischoff KJ, Geiger AM. Diffusion of aromatase inhibitors for breast cancer therapy between 1996 and 2003 in the Cancer Research Network. Breast Cancer Res Treat. 2008 Feb;107(3):397-403. PMID: 17393300.

Purpose: Clinical trials demonstrated adjuvant aromatase inhibitor treatment is superior for decreasing breast cancer recurrence risk over adjuvant tamoxifen treatment as early as 2P001. Yet clinical use for adjuvant treatment was not recommended by the American Society of Clinical Oncology until 2004. Aromatase inhibitor uptake after the first public presentation of randomized trial results but before the release of national guidelines is unclear. We evaluated diffusion of aromatase inhibitor dispensings for breast cancer treatment in integrated healthcare delivery systems across the United States.
Methods: We collected automated data for 13,245 women enrolled at seven integrated healthcare delivery systems in the Cancer Research Network. All women were aged >55 and diagnosed with estrogen receptor positive, invasive breast

cancer between 1996 and 2003. We used electronic pharmacy data to identify aromatase inhibitor and tamoxifen dispensings through 2004. We evaluated the proportions of women who received hormone dispensings in two ways: (1) at any point after diagnosis to capture all use, and (2) in the two-year period following diagnosis to approximate adjuvant use.

Results: *Over time, adjuvant aromatase inhibitor use increased whereas tamoxifen use decreased. Aromatase inhibitor dispensings within 2 years of diagnosis increased from 4.1% among women diagnosed in 2000 to 13% in 2001, 24% in 2002, and 40% in 2003. Tamoxifen use declined starting in 2001 at every system.*

Conclusion: *Aromatase inhibitor use rose dramatically after 2001 while tamoxifen use decreased. It appears results from early clinical trials changed practice in these integrated healthcare systems before formal changes in national guidelines.*

70. Brown JS, Kulldorff M, Petronis KR, Reynolds R, Chan KA, Davis RL, Graham D, Andrade SE, Raebel MA, Herrinton L, Roblin D, Boudreau D, Smith D, Gurwitz JH, Gunter MJ, Platt R. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. Pharmacoepidemiol Drug Saf. 2009 Mar;18(3):226-34. PMID: 19148879.

Purpose: *Active surveillance of population-based health networks may improve the timeliness of detection of adverse events (AEs). Our objective was to expand our previous signal detection work by investigating the effect on signal detection of alternative study specifications.*

Methods: *We compared the signal detection performance under various study specifications using historical data from nine health plans involved in the HMO Research Network's Center for Education and Research on Therapeutics (CERT). Five drug-event pairs representing generally accepted associations with an AE and two pairs representing "negative controls" were analyzed. Alternative study specifications related to the definition of incident users and incident AEs were assessed and compared to our previous findings.*

Results: *Relaxing the incident AE exclusion criteria by (1) including members with prior outpatient diagnoses of interest and (2) halving (to 90 days) the time window specified to define incident exposure and diagnoses increased the number of members under surveillance and as a consequence increased the number of exposed days and diagnoses by about 10-20%. The alternative specifications tend to result in earlier signal detection by 10-16 months, a likely consequence of more exposures and events entering the analysis.*

Conclusions: *This paper provides additional preliminary information related to conducting prospective safety monitoring using health plan data and sequential analytic methods. Our findings support continued investigation of using health plan data and sequential analytic methods as a potentially important contribution to active drug safety surveillance.*

71. Stopponi MA, Alexander GL, McClure JB, Carroll NM, Divine GW, Calvi JH, Rolnick SJ, Strecher VJ, Johnson CC., et al. Recruitment to a randomized web-based nutritional intervention trial: characteristics of participants compared to non-participants. *J Med Internet Res* 2009;11:e38. PMID: 19709990.

Background: *Web-based behavioral programs efficiently disseminate health information to a broad population, and online tailoring may increase their effectiveness. While the number of Internet-based behavioral interventions has*

grown in the last several years, additional information is needed to understand the characteristics of subjects who enroll in these interventions, relative to those subjects who are invited to enroll.

Objective: *The aim of the study was to compare the characteristics of participants who enrolled in an online dietary intervention trial (MENU) with those who were invited but chose not to participate, in order to better understand how these groups differ.*

Methods: *The MENU trial was conducted among five health plans participating in the HMO Cancer Research Network in collaboration with the University of Michigan Center for Health Communication Research. Approximately 6000 health plan members per site, between the ages of 21 and 65, and stratified by gender with oversampling of minority populations, were randomly selected for recruitment and were mailed an invitation letter containing website information and a US$2 bill with the promise of US$20 for completing follow-up surveys. Administrative and area-based data using geocoding along with baseline survey data were used to compare invitees (HMO members sent the introductory letter), responders (those who entered a study ID on the website), and enrollees (those who completed the enrollment process). Generalized estimating equation multivariate and logistic regression models were used to assess predictors of response and enrollment.*

Results: *Of 28,460 members invited to participate, 4270 (15.0%) accessed the website. Of the eligible responders, 2540 (8.9%) completed the consent form and baseline survey and were enrolled and randomized. The odds of responding were 10% lower for every decade of increased age (P < .001), while the likelihood of enrolling was 10% higher for every decade increase in age (P < .001). Women were more likely to respond and to enroll (P < .001). Those living in a census tract associated with higher education levels were more likely to respond and enroll, as well as those residing in tracts with higher income (P < .001). With a 22% (n = 566) enrollment rate for African Americans and 8% (n = 192) for Hispanics, the enrolled sample was more racially and ethnically diverse than the background sampling frame.*

Conclusions: *Relative to members invited to participate in the Internet-based intervention, those who enrolled were more likely to be older and live in census tracts associated with higher socioeconomic status. While oversampling of minority health plan members generated an enrolled sample that was more racially and ethnically diverse than the overall health plan population, additional research is needed to better understand methods that will expand the penetration of Internet interventions into more socioeconomically diverse populations.*

72. Hazlehurst B, Naleway A, Mullooly J. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. *Vaccine* 2009;27:2077-2083. PMID: 19428833.

The Vaccine Safety Datalink (VSD) is a collaboration between the CDC and eight large HMOs to investigate adverse events following immunization through analyses of clinical data. We modified an existing system, called MediClass, that uses natural language processing to identify clinical events recorded in electronic medical records (EMRs). We customized MediClass so it could detect possible vaccine adverse events (VAEs) generally, and gastrointestinal-related VAEs in particular, in the text clinical notes of encounters recorded in the EMR of a large HMO. Compared to methods that use diagnosis and utilization codes assigned to encounters by

clinicians and administrators, the MediClass system can both find more adverse events and improve the positive predictive value for detecting possible VAEs.

73. Yih WK, Kulldorff M, Fireman BH, Shui IM, Lewis EM, Klein NP, Baggs J, Weintraub ES, Belongia EA., et al. Active surveillance for adverse events: the experience of the Vaccine Safety Datalink project. *Pediatrics* 2011;127 Suppl 1:S54-S64. PMID: 21502252.

    Objective: *To describe the Vaccine Safety Datalink (VSD) project's experience with population-based, active surveillance for vaccine safety and draw lessons that may be useful for similar efforts.*
    Patients and Methods: *The VSD comprises a population of 9.2 million people annually in 8 geographically diverse US health care organizations. Data on vaccinations and diagnoses are updated and extracted weekly. The safety of 5 vaccines was monitored, each with 5 to 7 pre-specified outcomes. With sequential analytic methods, the number of cases of each outcome was compared with the number of cases observed in a comparison group or the number expected on the basis of background rates. If the test statistic exceeded a threshold, it was a signal of a possible vaccine-safety problem. Signals were investigated by using temporal scan statistics and analyses such as logistic regression.*
    Results: *Ten signals appeared over 3 years of surveillance: 1 signal was reported to external stakeholders and ultimately led to a change in national vaccination policy, and 9 signals were found to be spurious after rigorous internal investigation. Causes of spurious signals included imprecision in estimated background rates, changes in true incidence or coding over time, other confounding, inappropriate comparison groups, miscoding of outcomes in electronic medical records, and chance. In the absence of signals, estimates of adverse-event rates, relative risks, and attributable risks from up-to-date VSD data have provided rapid assessment of vaccine safety to policy-makers when concerns about a specific vaccine have arisen elsewhere.*
    Conclusions: *Care with data quality, outcome definitions, comparison groups, and length of surveillance are required to enable detection of true safety problems while minimizing false signals. Some causes of false signals in the VSD system were preventable and have been corrected, whereas others will be unavoidable in any active surveillance system. Temporal scan statistics, analyses to control for confounding, and chart review are indispensable tools in signal investigation. The VSD's experience may inform new systems for active safety surveillance.*

74. Greene SK, Kulldorff M, Yin R, Yih WK, Lieu TA, Weintraub ES, Lee GM. Near real-time vaccine safety surveillance with partially accrued data. *Pharmacoepidemiol Drug Saf* 2011;20:583-590. PMID: 21538670.

    Purpose: *The Vaccine Safety Datalink (VSD) Project conducts near real-time vaccine safety surveillance using sequential analytic methods. Timely surveillance is critical in identifying potential safety problems and preventing additional exposure before most vaccines are administered. For vaccines that are administered during a short period, such as influenza vaccines, timeliness can be improved by undertaking analyses while risk windows following vaccination are ongoing and by accommodating predictable and unpredictable data accrual delays. We describe practical solutions to these challenges, which were adopted by the VSD Project during pandemic and seasonal influenza vaccine safety surveillance in 2009/2010.*

Methods: Adjustments were made to two sequential analytic approaches. The Poisson-based approach compared the number of pre-defined adverse events observed following vaccination with the number expected using historical data. The expected number was adjusted for the proportion of the risk window elapsed and the proportion of inpatient data estimated to have accrued. The binomial-based approach used a self-controlled design, comparing the observed numbers of events in risk versus comparison windows. Events were included in analysis only if they occurred during a week that had already passed for both windows.
Results: Analyzing data before risk windows fully elapsed improved the timeliness of safety surveillance. Adjustments for data accrual lags were tailored to each data source and avoided biasing analyses away from detecting a potential safety problem, particularly early during surveillance.
Conclusions: The timeliness of vaccine and drug safety surveillance can be improved by properly accounting for partially elapsed windows and data accrual delays.

75. McCarthy NL, Gee J, Weintraub E, Donahue JG, Nordin JD, Daley MF, Naleway A, Henninger M, Baxter R., et al. Monitoring vaccine safety using the Vaccine Safety Datalink: utilizing immunization registries for pandemic influenza. *Vaccine* 2011;29:4891-4896. PMID: 21596088.

Mass vaccination campaigns during which new vaccines may be administered to many millions of people in a short period of time call for timely and accurate post-licensure surveillance to monitor vaccine safety. To address the need for timely H1N1 influenza vaccine safety information during the 2009-2010 H1N1 influenza pandemic, the Vaccine Safety Datalink (VSD) project assessed the feasibility and potential mechanisms for utilizing data from state and local immunization registries to capture vaccinations that would not otherwise be captured by the data systems of the participating VSD managed care organizations (MCOs). Three of the eight VSD sites were able to capture H1N1 immunization data electronically from the state and local registries, and one site was able to capture the immunizations through a paper-based system; however, the remaining four sites encountered various obstacles that prevented capture of such data. Additional work will be required at these sites to overcome the barriers, which included privacy and confidentiality laws, time constraints brought on by the pandemic, as well as data quality concerns.

76. Koebnick C, Langer-Gould AM, Gould MK, Chao CR, Iyer RL, Smith N, Chen W, Jacobsen SJ. Sociodemographic Characteristics of Members of a Large, Integrated Health Care System: Comparison with US Census Bureau Data. Perm J 2012 Summer;16(3):37-41. PMID: 23012597.

Background: Data from the memberships of large, integrated health care systems can be valuable for clinical, epidemiologic, and health services research, but a potential selection bias may threaten the inference to the population of interest.
Methods: We reviewed administrative records of members of Kaiser Permanente Southern California (KPSC) in 2000 and 2010, and we compared their sociodemographic characteristics with those of the underlying population in the coverage area on the basis of US Census Bureau data.
Results: We identified 3,328,579 KPSC members in 2000 and 3,357,959 KPSC members in 2010, representing approximately 16% of the population in the coverage area. The distribution of sex and age of KPSC members appeared to be similar to

the census reference population in 2000 and 2010 except with a slightly higher proportion of 40 to 64 year olds. The proportion of Hispanics/Latinos was comparable between KPSC and the census reference population (37.5% vs 38.2%, respectively, in 2000 and 45.2% vs 43.3% in 2010). However, KPSC members included more blacks (14.9% vs 7.0% in 2000 and 10.8% vs 6.5% in 2010). Neighborhood educational levels and neighborhood household incomes were generally similar between KPSC members and the census reference population, but with a marginal underrepresentation of individuals with extremely low income and high education.
Conclusions: *The membership of KPSC reflects the socioeconomic diversity of the Southern California census population, suggesting that findings from this setting may provide valid inference for clinical, epidemiologic, and health services research.*

77. Desai JR, Wu P, Nichols GA, Lieu TA, O'Connor PJ. Diabetes and asthma case identification, validation, and representativeness when using electronic health data to construct registries for comparative effectiveness and epidemiologic research. Med Care. 2012 Jul;50 Suppl:S30-5. PMID: 22692256.

Background: *Advances in health information technology and widespread use of electronic health data offer new opportunities for development of large scale multisite disease-specific patient registries. Such registries use existing data, can be constructed at relatively low cost, include large numbers of patients, and once created can be used to address many issues with a short time between posing a question and obtaining an answer. Potential applications include comparative effectiveness research, public health surveillance, mapping and improving quality of clinical care, and others.*
Objective and Discussion: *This paper describes selected conceptual and practical challenges related to development of multisite diabetes and asthma registries, including development of case definitions, validation of case identification methods, variation in electronic health data sources; representativeness of registry populations, including the impact of attrition. Specific challenges are illustrated with data from actual registries.*

VII.   Administrative Processes in Research

78. Greene SM, Hart G, Brown M, Wagner E. Measuring and Improving Performance in Multi-Center Research Consortia. J Natl Cancer Inst Monograph 2005 Nov;35(3):26-32. PMID: 16287882.

Background: *Some evidence suggests that the quality of the organization and management of research consortia influences productivity and staff satisfaction. Collaborators in a research consortium generally focus on developing and implementing studies and thus rarely assess the process of collaboration. We present an approach to evaluating and improving a research consortium, using the HMO Cancer Research Network (CRN) as an example.*
Methods: *Five domains are evaluated: extent of collaboration and quality of communication; performance of projects and infrastructure; data quality; scientific productivity; and impact on member organizations. The primary assessment tool is a survey of CRN scientists and project staff, undertaken annually.*
Results: *Each year, the evaluation has identified critical aspects of this collaboration that could be improved. Several tangible changes have been implemented to improve productivity of the consortium. The most important result of the CRN*

Evaluation is the ability to have open dialogue about ways to improve its overall performance.
Conclusion: *Optimizing the process of collaboration will contribute to achievement of the scientific goals. The experience of the CRN provides a useful framework and process for evaluating the structure of consortium-based* research.

79. Greene SM, Larson EB, Boudreau DM, Johnson KE, Ralston J, Reid R, Fishman P. The coordinated clinical studies network: a multidisciplinary alliance to facilitate research and improve care. Perm J. 2005 Fall;9(4):33-5. PMID: 22811643.

    The NIH Roadmap is a major effort to reshape the US health research enterprise to accelerate medical discovery and to do so in such a way that actually hastens population health improvement through research. The Roadmap's ultimate goal resonates with the HMO Research Network, a consortium of integrated health care systems that uses its collective scientific capabilities to integrate research, practice, and policy for the improvement of health and health care among diverse populations. (See page 6 for abstracts from the HMO Research Network annual conference.) As such, the HMO Research Network was ideally suited to propose a new consortium project as a part of the NIH Roadmap, the Coordinated Clinical Studies Network (CCSN). The CCSN was funded in 2004 to create a path-breaking research facility that leverages several distinctive features of the HMO Research Network: the multidisciplinary scientific capabilities of its researchers; the ability to rapidly move clinical research findings into care delivery; its large, diverse patient populations; and a commitment to placing its findings in the public domain. Among the goals of the CCSN are to augment the capacity and infrastructure for conducting research, and to use considerable investments in health informatics to improve the scope and efficiency of research data collection. The NIH Roadmap is a revolutionary step toward a new paradigm for research and responds to both a compelling social need and rapid technological advances in biomedicine. The CCSN's participation in the Roadmap Initiative is a unique opportunity for researchers, clinicians, and our patients.

80. Williams RL, Johnson SB, Greene SM, Larson EB, Green LA, Morris A, Confer D, Reaman G, Madigan R, Kahn J; Principal Investigators of Clinical Research Networks Initiative. Signposts along the NIH roadmap for reengineering clinical research: lessons from the Clinical Research Networks initiative. Arch Intern Med. 2008 Sep 22;168(17):1919-25. PMID: 18809820.

    Background: *The National Institutes of Health (NIH) Roadmap for Medical Research aims to increase the efficiency and speed of clinical research. We report results and lessons learned from a key component of the Roadmap, the Clinical Research Networks initiative.*
    Methods: *Twelve diverse, experienced, large, clinical research networks were funded for 3 years to develop strategies for integrating, expanding, and increasing the interoperability of clinical research networks in support of the Roadmap goals. Network leaders met periodically in person and by teleconference to describe common challenges encountered and solutions used for expansion and increased interoperability.*
    Results: *These networks developed innovative solutions to technical challenges, including strategies for interoperability of information systems and management of complex information system technologies (e.g., "brokering" to address data system incompatibility, data transfer, and security requirements), and solutions to human*

factor challenges at the individual, group, intraorganizational, and interorganizational levels (eg, applying collaborative organizing and decision-making processes based on key principles).

Conclusions: *These solutions can provide guidance to existing and future clinical research networks, particularly those forming as part of the NIH Clinical Translation Science Award program. Remaining technical and human factor challenges, however, as well as the largely unmet need for consistent funding for network infrastructure and maintenance, stand in the way of fulfilling the vision of a robust future role for clinical research networks.*

VIII.   Knowledge Management and Resources

81. Dolor RJ, Greene SM, Thompson E, Baldwin LM, Neale AV. Partnership-driven Resources to Improve and Enhance Research (PRIMER): a survey of community-engaged researchers and creation of an online toolkit. Clin Transl Sci. 2011 Aug;4(4):259-65. PMID: 21884512.

Objective: *This project aimed to develop an open-access website providing adaptable resources to facilitate best practices for multisite research from initiation to closeout.*
Methods: *A web-based assessment was sent to the leadership of the Clinical and Translational Science Award (CTSA) Community Engagement Key Functions Committee (n= 38) and the CTSA-affiliated Primary Care Practice-based Research Networks (PBRN, n= 55). Respondents rated the benefits and barriers of multisite research, the utility of available resources, and indicated their level of interest in unavailable resources. Then, existing research resources were evaluated for relevance to multisite research, adaptability to other projects, and source credibility.*
Results: *Fifty-five (59%) of invited participants completed the survey. Top perceived benefits of multisite research were the ability to conduct community-relevant research through academic-community partnerships (34%) and accelerating translation of research into practice (31%). Top perceived barriers were lack of research infrastructure to support PBRNs and community partners (31%) and inadequate funding to support multisite collaborations (26%). Over 200 resources were evaluated, of which 120 unique resources were included in the website.*
Conclusion: *The PRIMER Research Toolkit (http://www.researchtoolkit.org) provides an array of peer-reviewed resources.*

82. Greene SM, Baldwin LM, Dolor RJ, Thompson E, Neale AV. Streamlining research by using existing tools. Clin Transl Sci. 2011 Aug;4(4):266-7. PMID: 21884513.

Over the past two decades, the health research enterprise has matured rapidly, and many recognize an urgent need to translate pertinent research results into practice, to help improve the quality, accessibility, and affordability of US healthcare. Streamlining research operations would speed translation, particularly for multisite collaborations. However, the culture of research discourages reusing or adapting existing resources or study materials. Too often, researchers start studies and multisite collaborations from scratch-reinventing the wheel. Our team developed a compendium of resources to address inefficiencies and researchers' unmet needs and compiled them in a research toolkit website (http://www.ResearchToolkit.org). Through our work, we identified philosophical and operational issues related to

disseminating the tool kit to the research community. We explore these issues here, with implications for the nation's investment in biomedical research.

83. Websites

Research in the HMO Research Network: Research Process and Partnership Primer
http://www.hmoresearchnetwork.org/resources/toolkit/HMORN_Research-Process-and-Partnership-Primer.pdf

General Introduction to the HMORN

o   Organizational and Operational Orientation to the HMORN
     http://www.hmoresearchnetwork.org/resources/tools/HMORN_Operational-Orientation.ppt

o   HMORN Precis
     http://www.hmoresearchnetwork.org/resources/tools/HMORN_Precis.pdf

Guidance for Multi-Center research
o   Collaboration Toolkit:
     http://www.hmoresearchnetwork.org/resources/toolkit/HMORN_CollaborationToolkit.pdf

Virtual Data Warehouse (VDW)
o   Questions and Answers About the Virtual Data Warehouse
     http://www.hmoresearchnetwork.org/resources/tools/HMORN_VDW-Questions-and-Answers.pdf

SubAward Templates
o   SubAward Template:
     http://www.hmoresearchnetwork.org/resources/toolkit/HMORN_SubAward_Template.doc

Data Use Agreements (DUAs)

o   DUA Template:
     http://www.hmoresearchnetwork.org/resources/toolkit/HMORN_DUA_Template.doc

o   Reciprocal DUA Template:
     http://www.hmoresearchnetwork.org/resources/toolkit/HMORN_Reciprocal_DUA_Examples.doc

Policies/Procedures for Multi-Center project closeout
o   Multi-Center Closeout Guide
     http://www.hmoresearchnetwork.org/resources/toolkit/HMORN_CloseoutGuide.pdf

IX. HMORN and Conference Summaries

84. Greenlee RT, Coleman LA, Nelson AF, Selby JV. Partnerships in Translation: Advancing Research and Clinical Care. The 14th Annual HMO Research Network Conference, April 13-16, 2008, Minneapolis, Minnesota. Clin Med Res. 2008 Dec;6(3-4):109-12. PMID: 19325174.

    The Health Maintenance Organization Research Network held its annual meeting in Minneapolis in April of 2008, with more than 300 investigators, research staff, clinical leaders, and academic partners gathering in conjunction with the conference theme 'Partnerships in Translation: Advancing Research and Clinical Care.' This article provides some background on the network, its research activities, and the annual conference. Also featured is an article by Coleman and colleagues summarizing the conference's first plenary session, where operational leaders of health care organizations discussed the optimization of health care through research. This issue of Clinical Medicine & Research also includes a selection of scientific abstracts presented at the meeting on a wide range of clinical and population health topics.

85. Shah NR, Stewart WF. Clinical Effectiveness: Leadership in Comparative Effectiveness and Translational Research. The 15[Th] Annual HMO Research Network Conference, April 26-29, 2009, Danville, Pennsylvania. *Clin Med Res.* 2010 8:28-29. PMID: 20305148.

    The Health Maintenance Organization Research Network (HMORN), a consortium of 16 health care delivery systems with integrated research divisions, held its annual meeting in Danville, Pennsylvania in April of 2009 and was attended by more than 260 researchers and operational leaders from HMORN organizations, pharmaceutical companies, the National Cancer Institute, and the Agency for Healthcare Research and Quality. The 2009 meeting was held from April 26th to April 29th at the Henry Hood Center for Health Research, and was hosted by Geisinger Health System. The conference theme was "Clinical Effectiveness: Leadership in Comparative Effectiveness and Translational Research." This article provides some background on the network, its research activities, and the annual conference. This issue of Clinical Medicine & Research also includes selected scientific abstracts presented at the meeting.

86. Stevens AB, Sanghi S. Emerging Frontiers in Healthcare Research and Delivery. The 16th Annual HMO Research Network Conference, March 21-24, 2010, Austin, Texas. *Clin Med Res.* 2010 Dec;8(3-4):176-8. PMID: 21177536.

    The Health Maintenance Organization Research Network (HMORN), a consortium of 16 healthcare delivery systems with integrated research centers, held their 16th annual conference in Austin, Texas from March 21-24, 2010. The conference was hosted by Scott & White Healthcare. Its theme "Emerging Frontiers in Healthcare Research and Delivery" reflected the objective of the conference which was to build synergy among scientists and clinicians to influence the health of the nation; to demonstrate the network's commitment to reach beyond traditional collaborators; discuss tools and technologies; and to expand opportunities for public-private partnerships in cutting-edge healthcare research and delivery. More than 320 researchers and healthcare professionals, representing each of the member HMOs, participated in this conference.

    Representatives from the AHRQ, CDC, NCI and NIH met with researchers to advance the quality and breadth of public domain research in HMOs. The objective

of this article is to provide information about the HMORN and its 16th annual conference.

87. Lieu TA, Hinrichsen, VL, Moreira A, Platt R. Collaborations in Population-Based Health Research. The 17th Annual HMO Research Network Conference, March 23-25, 2011, Boston, Massachusetts. *Clin Med Res.* 2011; 9, No. 3/4: 137-140. PMID: 22090515.

The HMO Research Network (HMORN) is a consortium of 16 health care systems with integrated research centers. Approximately 475 people participated in its 17th annual conference, hosted by the Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School. The theme, "Collaborations in Population-Based Health Research," reflected the network's emphasis on collaborative studies both among its members and with external investigators. Plenary talks highlighted the initial phase of the HMORN's work to establish the NIH-HMO Collaboratory, opportunities for public health collaborations, the work of early career investigators, and the state of the network. Platform and poster presentations showcased a broad spectrum of innovative public domain research in areas including disease epidemiology and treatment, health economics, and information technology. Special interest group sessions and ancillary meetings provided venues for informal conversation and structured work among ongoing groups, including networks in cancer, cardiovascular diseases, lung diseases, medical product safety, and mental health.

88. Newton, KM, Larson E.B. Learning Health Care Systems: Leading Through Research, The 18th Annual HMO Research Network Conference, April 29-May 2, 2012, Seattle, Washington. *Clin Med Res.* 2012, 10:140-142. PMID: 22904375.

The Health Maintenance Organization Research Network (HMORN), a consortium of 19 health care delivery systems with integrated research centers, held their 18th annual conference in Seattle, Washington from April 29 to May 2, 2012. Group Health Research Institute hosted the conference, "Learning health care systems: Leading through research". The 2012 theme was chosen to reflect the critical role of collaboration among researchers, clinicians and health systems to improve health care nationally. Over 500 researchers and health care professionals participated in this conference. Representatives from the NCI, NHLBI, NIMH, and PCORI met with researchers to advance the quality and breadth of public domain research in HMOs. In this article we summarize information about the HMORN and its 18th annual conference.

89. Copeland LA, and Zeber, JE. Advancing Research in the Era of Healtcare Reform: The 19th Annual HMO Research Network Conference, April 16-18, 2013, San Francisco, California. Clin Med Res. 2013,11:120-122. PMCID: PMC3788495.

Healthcare system transformation has pursued higher quality, more affordable healthcare from the popularization of health maintenance organizations (HMOs) in the 1970s, through the innovations in structure and management of the 1990s, to the current-day challenge of extending healthcare coverage to previously uninsured groups as partially conceptualized in the Patient Protection and Affordable Care Act (PPACA) toward Accountable Care Organizations (ACOs). Census 2010 reported that 17% of Americans had no insurance whatsoever, 18% were publicly insured,1 and the remaining 65% were privately insured. Overall, approximately 1 in 4 Americans is enrolled in an HMO, including Medicare HMO options. Research on

persons covered by managed care insurance and integrated delivery systems is needed to round out our understanding of how people use healthcare and how their health status relates to that usage. Now in its 19th year, the HMORN held its annual meeting in San Francisco, April 16–18, 2013, hosted by Kaiser Permanente of Northern California with Dr. Alan Go as Chair.