

# DIGITAL DATA PRIORITIES FOR CONTINUOUS LEARNING IN HEALTH AND HEALTH CARE

An Institute of Medicine Workshop sponsored by the Office of the National Coordinator for Health Information Technology



March 23, 2012 Keck Center, The National Academies 500 Fifth Street, NW Washington, DC 20001



# Table of Contents

# SECTION 1: WORKSHOP FRAMING MATERIALS

- Agenda
- Planning Committee Roster
- Participant List

# SECTION 2: BACKGROUND MATERIALS

- Ancker, Jessica S et al. Root Causes Underlying Challenges to Secondary Use of Data. AMIA Annual Symposium Proceedings. 2011.
- Berhman, Rachel E. et al. Developing the Sentinel System A National Resource for Evidence Development. The New England Journal of Medicine. 2011.
- Brownstein, Jason S. et al. Rapid Identification of Myocardial Infarction Risk Associated With Diabetes Medications Using Electronic Medical Records. Diabetes Care. 2010.
- Buck, Michael D. et al. *The Hub Population Health System: Distributed ad hoc Queries and Alerts.* Journal of the American Medical Informatics Association. 2011.
- Dixon, Brian E. et al. *Electronic Laboratory Data Quality and the Value of a Health Information Exchange to Support Public Health Reporting Processes.* AMIA Annual Symposium Proceedings. 2011.
- Mathias, Jason S. Use of Electronic Health Record Data to Evaluate Overuse of Cervical Cancer Screening. Journal of the American Medical Informatics Association. 2012.
- Parsons, Amanda et al. *Validity of Electronic Health Record-Derived Quality Measurement for Performance Monitoring*. Journal of the American Medical Informatics Association. 2012.
- Platt, Richard et al. *The U.S. Food and Drug Administration's Mini-Sentinel program: Status and Direction.* Pharmacoepidemiology and Drug Safety. 2012.
- Tannen, Richard L. et al. Use of Primary Care Electronic Medical Record Database in Drug Efficacy Research on Cardiovascular Outcomes: Comparison of Database and Randomised Controlled Trial Findings. British Journal of Medicine. 2009.
- U.S. Department of Health and Human Services. *Guidance for Industry Part 11, Electronic Records; Electronic Signatures Scope and Application.* 2003.
- U.S. Department of Health and Human Services. *Code of Federal Regulations: Title 21, Volume 1.* 2011.
- VSRT Staff. Data Analysis Inventory. 2012.

# SECTION 3: BIOGRAPHIES AND MEETING LOGISTICS

- Speaker Biographies
- Location, Hotel, and Travel

# Workshop Framing Materials



# DIGITAL DATA PRIORITIES FOR CONTINUOUS LEARNING IN HEALTH AND HEALTH CARE

An Institute of Medicine Workshop Sponsored by the Office of the National Coordinator for Health Information Technology

MARCH 23, 2011

KECK CENTER 500 FIFTH STREET NW, WASHINGTON DC 20001

# A LEARNING HEALTH SYSTEM ACTIVITY IOM ROUNDTABLE ON VALUE & SCIENCE-DRIVEN HEALTH CARE

# Meeting objectives

- 1. Discuss the current quality status of digital health data.
- 2. Explore challenges, and identify key questions related to data quality in the use of EHRs, patient registries, administrative data, and public health sources for learning—continuous and episodic—and for system operational and improvement purposes.
- 3. Engage individuals and organizations leading the way in improving the reliability, availability, and usability of digital health data for real-time knowledge generation and health improvement in a continuously learning health system.
- 4. Identify and characterize the current deficiencies and consider strategies, priorities, and responsibilities to address the deficiencies.
- 5. Initiate the development of a strategic framework for integrated and networked stewardship of efforts to continuously increase digital data utility.

# Agenda

7:30 am Coffee and light breakfast available

8:00 am	Welcome, introductions and overview
	<ul> <li>Welcome, framing of the meeting and agenda overview</li> <li>Michael McGinnis (Institute of Medicine)</li> <li>Farzad Mostashari (Office of the National Coordinator)</li> <li>Jim Walker (Planning Committee Chair)</li> </ul>
8:15 am	Characteristics, challenges, and determinants of data quality
	Session Description: This session includes brief comments on the data quality challenges that lie ahead and a longer discussion of the characteristics and determinants of digital health data quality.

# ➢ Key Topics:

- Challenges on the horizon Doug Fridsma (ONC)
- Characteristics and determinants of data quality *Marc Overhage (Siemens)*

# **OPEN DISCUSSION**

9:00am	Performance assessment
	Session Description: This session focuses on the quality of digital health
	data needed to evaluate clinical care delivery, population management and the business and operating processes that make up a learning health system

# ➢ Key Topics:

• Assessing value Carol McCall (GNS)

• Managing populations and processes Mark Leenay (OptumHealth)

# **OPEN DISCUSSION**

10:00am	Break
10:15am	Enabling research
	Session Description: This session focuses on the quality of digital health data needed to enable research.
	<ul> <li>Key Topics:</li> <li>Clinical research Rebecca Kush (Clinical Data Interchange Standards Consortium)</li> </ul>
	• Translational informatics Mia Levy (Vanderbilt)
	OPEN DISCUSSION
11:15am	Supporting public health and surveillance

- Session Description: This session focuses on the quality of digital health data needed to support of public health functions, including surveillance
- ➢ Key Topics:
  - Public health surveillance and management *James Buehler (CDC)*
  - State-level perspective Martin LaVenture (Minnesota Dept of Health)

# **OPEN DISCUSSION**

12:15pm	Lunch keynote
	Who is your customer?

James Heywood, PatientsLikeMe

1:00pm	Approaches to continuous improvement using large-scale data sets
	Session Description: Session presentations will focus on the implications of digital health data quality on the potential for learning from large amounts of health data.
	Key Topics:
	• Using distributed data/ Query Health Rich Platt (Harvard) and Rich Elmore (ONC)
	<ul> <li>Data analysis and discovery of significant patterns David Madigan (OMOP/Columbia)</li> </ul>
	OPEN DISCUSSION
2:00pm	Innovative approaches to addressing data challenges
	Session Description: This session will focus on innovative approaches to overcoming some prominent challenges associated with using health data.
	> Topics:

- Data harmonization Chris Chute (Mayo)
- Linking data across time and sources Vik Kheterpal (CareEvolution Inc)

# **OPEN DISCUSSION**

# 3:00pm Strategies going forward

- Session Description: This session will include a rapid-fire, moderated discussion to identify the top 10 actions necessary for progress discussed during the course of the meeting.
- 1. Identification of potential action steps 20 min. (45 seconds each)
- 2. Rapid identification of pros and cons 15 min.
- 3. Identification of top ten leading action steps 25 min.

# **OPEN DISCUSSION**

4:00pm	Next steps

Session Description: This session will build off of the ten action steps identified in the previous session and outline options to move forward.

5:00pm Adjourn

Planning Committee	
James Walker (chair)	Geisinger Health System
Justine Carr	Steward Health Care
William DuMouchel	Oracle
Jamie Heywood	Patients Like Me
Rebecca D. Kush	Clinical Data Standards Interchange Consortium
Lisa M. Lee	Centers for Disease Control and Prevention
Theresa M. Mullin	Food and Drug Administration
Lucila Ohno-Machado	University of California San Diego
Richard Platt	Harvard University
Jim Scanlon	U.S. Department of Health and Human Services
Paul Stang	Johnson & Johnson
Walter Suarez	Kaiser Permanente



# DIGITAL DATA PRIORITIES FOR CONTINUOUS LEARNING IN HEALTH AND HEALTH CARE

Workshop Planning Committee

# <u>Chair</u>

James Walker, MD Chief Information Officer Geisinger Health System

# Members

**Justine Carr, MD** Chief Medical Officer Steward Health Care

**William DuMouchel, PhD** Chief Statistical Scientist Oracle Health Sciences

**Jamie Heywood** Chairman Patients Like Me

**Rebecca D. Kush, PhD** President and CEO Clinical Data Standards Interchange Consortium

Lisa M. Lee, PhD Chief Science Officer CDC (OELS)

**Theresa M. Mullin, PhD** Director, Office of Planning and Informatics FDA (CDER) **Lucila Ohno-Machado, MD, PhD** Founding Chief, Biomedical Informatics University of California San Diego

**Richard Platt, MD, MS** Chair, Population Medicine Harvard University

**Jim Scanlon** Deputy ASPE (Science and Policy) Department of Health and Human Services

Paul Stang, PhD Senior Director of Epidemeology Johnson & Johnson

Walter Suarez, MD, MPH Director, Health IT Strategy and Policy Kaiser Permanente



# Digital Data Priorities for Continuous Learning in Health and Health Care March 23, 2012

# Workshop Participants

Sameer Bade, MD SM Director, Clinical Solutions Microsoft Health Solutions Group

Jesse Berlin, ScD Vice President, Epidemiology Janssen Research & Development

Hunt Blair State Health IT Coordinator Department of Vermont Health Access

Kathleen Blake, MD, MPH Senior Research Director Center for Medical Technology Policy

Meryl Bloomrosen, MBA Vice President, Public Policy and Government Relations American Medical Informatics Association

William B. Borden, MD Senior Advisor, Office of Health Policy U.S. Department of Health and Human Services

Jeffrey Brown, PhD Assistant Professor, Population Medicine Harvard Medical School

James W. Buehler, MD Director, Public Health Surveillance Centers for Disease Control and Prevention

Ken H. Buetow, PhD Associate Director, Bioinformatics and IT National Cancer Institute Darby Buroker Director, Beacon Program Management Rhode Island Quality Institute

Justine Carr, MD Chief Medical Officer Steward Health Care

Basit Chaudhry, MD, PhD Senior Researcher, Clinical Transformation IBM Research

Christopher G. Chute, MD, DrPH, MPH Professor, Medical Informatics Mayo Clinic

David C. Classen, MD, MS Senior Partner Computer Sciences Corporation

Elaine Collier, MD Office of Policy, Communications, and Strategic Alliances National Center for Advancing Translational Sciences

Kevin Cranston, MDiv Director, Bureau of Infectious Disease Massachusetts Department of Public Health

Theresa Ann Cullen, MD, MS Director, Health Informatics Veterans Health Administration

Art Davidson, MD, MPH Director of Public Health Informatics Denver Public Health Department Donald E. Detmer, MD, MA, FACS Medical Director, Division of Advocacy and Health Policy American College of Surgeons

Gregory J. Downing, DO, PhD Project Director, Initiative on Personalized Health Care Department of Health and Human Services

William DuMouchel, PhD Chief Statistical Scientist, Health Sciences Oracle

Jim Dupree, MD, MPH Health Policy Fellow American College of Surgeons

Carl Dvorak Chief Operating Officer Epic Systems Corporation

Margaret Edmunds, PhD Senior Program Officer Institute of Medicine

Richard Elmore, MA Coordinator, Standards & Interoperability Office of the National Coordinator for HIT

Jeff Felton, MBA President RelayHealth

T. Bruce Ferguson, Jr., MD Chairman, Cardiovascular Sciences East Carolina University

Mark E. Fleury, PhD Associate Director, Science Policy American Association for Cancer Research

Doug Fridsma, MD, PhD Director, Office of Standards & Interoperability Office of the National Coordinator for HIT

Kathleen Gallagher, DSc, MPH Director, Notifiable Disease Surveillance and Healthcare Information, OELS Centers for Disease Control and Prevention Shaun J. Grannis, MD, MS, FAAFP Medical Informatics Researcher Regenstrief Institute

Carl Gunter, PhD Professor, Computer Science University of Illinois

Laura Hansen University of Virginia

Jamie Heywood Chairman PatientsLikeMe

George Hripcsak, MD, MS Vivian Beaumont Allen Professor of Biomedical Informatics Columbia University

Stanley M. Huff, MD Chief Medical Informatics Officer Intermountain Healthcare

Betsy L. Humphreys, MLS Deputy Director National Library of Medicine

Ashish Jha, MPH, MD Associate Professor Harvard Medical School

Swapna U. Karkare, MS Research Associate Center for Medical Technology Policy

Joy Keeler Tobin, MBA, FHIMSS Chief of Health Informatics, Center for Transforming Health MITRE Corporation

Rosemary Kennedy, RN, MBA Vice President, HIT National Quality Forum

Simon Kennedy Senior Partner and Managing Director Boston Consulting Group Vik Kheterpal, MD Principal CareEvolution

Dave Knutson Acting Director, Division of Health Care Quality and Outcomes Department of Health and Human Services

Clifford Ko, MD, FACS Medical Director, Division of Research and Optimal Patient Care American College of Surgeons

Joel Kupersmith, MD Chief Research and Development Officer Veterans Health Administration

Rebecca D. Kush, PhD President and CEO Clinical Data Standards Interchange Consortium

Steve Labkoff, MD, FACP Head of Strategic Programs AstraZeneca

Martin LaVenture, PhD Director, Center for Health Informatics and e-Health Minnesota Department of Health

Lisa M. Lee, PhD Chief Science Officer, OELS Centers for Disease Control and Prevention

Mark Leenay, MS, MD Chief Medical Officer and Senior Vice President OptumHealth Care Solutions

Mia Levy, MD, PhD Assistant Professor, Biomedical Informatics Vanderbilt University

Bettijoyce Lide HIT Scientific Advisor and Program Manager National Institutes of Standards & Technology Patricia MacTaggart, MBA, MMA Lead Research Scientist George Washington University

David Madigan, PhD Chair of Statistics Columbia University

Karen Matsuoka, DPhil, Mphil Research Director Brookings Institution

John Mattison, MD Chief Medical Information Officer Kaiser Permanente

Carol McCall, FSA, MAAA Strategy Officer GNS Healthcare

J. Alan Menius, Jr Senior Director, Medical Analytics GlaxoSmithKline R&D

Steven J. Meurer, PhD, MBA, MHS Senior Vice President, Comparative Data and Informatics University HealthSystem Consortium

Andrea Meyers, MPH Senior Pharmacoepidemiology Scientist Oracle Health Sciences

Blackford Middleton, MD, MPH, MSc Corporate Director of Clinical Informatics Research & Development Partners HealthCare

Rick Moore, FACHE Chief Information Officer National Committee for Quality Assurance

Farzad Mostashari, MD, MPH National Coordinator Office of the National Coordinator for HIT

Theresa M. Mullin, PhD Director, Office of Planning and Informatics, Center for Drug Evaluation and Research Food and Drug Administration Sharon B. Murphy, MD Senior Scholar Institute of Medicine

Judith Murphy, RN, FACMI, FAAN Deputy National Coordinator for Programs and Policy Office of the National Coordinator for HIT

Meredith Nahm, PhD Associate Director for Clinical Research Informatics Duke Translational Medicine Institute

Rachel Nelson, MHA Senior Advisor, Office of the National Coordinator Office of the National Coordinator for HIT

Lucila Ohno-Machado, MD, PhD Founding Chief, Division of Biomedical Informatics University of California San Diego

J. Marc Overhage, MD, PhD Chief Information Officer Siemens

Wilson Pace, MD Director, DARTNet University of Colorado

Richard Platt, MD, MS Chair, Population Medicine Harvard University

Daniel Pollock, MD Surveillance Branch Chief, Division of Healthcare Quality Promotion Centers for Disease Control and Prevention

Melissa Robb Project Director, Sentinel Initiative Food and Drug Administration Raj Sabharwal, MPH Senior Manager AcademyHealth

James V. Scanlon Deputy ASPE, Science and Data Policy U.S. Department of Health and Human Services

Jonathan C. Silverstein, MD, MS Vice President, Clinical Research Informatics NorthShore University HealthSystem

Paul Stang, PhD Senior Director of Epidemiology Johnson & Johnson

Walter Suarez, MD, MPH Director, Health IT Strategy and Policy Kaiser Permanente

Richard L. Tannen, MD Professor University of Pennsylvania

James Tcheng, MD Director, DTMI Biomedical Informatics Core Duke University Medical Center

Mark Tobias President Pantheon

Reed Tuckson, MD Executive VP & Chief of Medical Affairs UnitedHealth Group

Steven Waldren, MD Director, Center for Health IT American Academy of Family Physicians

James Walker, MD Chief Information Officer Geisinger Health System

# **Background Materials**

# **Root Causes Underlying Challenges to Secondary Use of Data**

Jessica S. Ancker, MPH, PhD,<sup>1,2</sup> Sarah Shih, MPH,<sup>3</sup> Mytri P. Singh, MPH,<sup>3</sup> Andrew Snyder, MPA,<sup>3</sup> Alison Edwards, MStat,<sup>1,2</sup> Rainu Kaushal, MD, MPH,<sup>1,2</sup> with the HITEC investigators

# <sup>1</sup>Division of Quality and Medical Informatics, Departments of Pediatrics and Public Health, Weill Cornell Medical College, New York, NY; <sup>2</sup>Health Information Technology Evaluation Collaborative (HITEC); <sup>3</sup>Primary Care Information Project (PCIP), New York City Department of Health and Mental Hygiene, New York, NY

### Abstract

Although one potential benefit of electronic information systems is the opportunity for secondary use of data, it is often challenging in practice to reuse data. We identify challenges to the secondary use of electronic data from a web-based project management system, and trace these challenges to their root causes. Data quality issues arose from: differential incentives for integrity of different data; software flexibility that allowed a single task to be documented in multiple ways; variability in documentation practices; variability in use of standardized vocabulary; and changes in project procedures and system configuration over time. These issues are very similar to the issues that pose challenges for secondary uses of clinical and operational data for research, public health, and quality improvement. We conclude that secondary use of operational data requires an in-depth understanding of the primary workflow processes that produced the data, as these processes lead to data integrity issues.

### Introduction

Although electronic information systems are developed to assist in day-to-day operational tasks, they offer the additional promise of data reuse for secondary purposes such as research, quality improvement, and public health.<sup>1-7</sup> Ambulance dispatch calls, retail pharmacy sales of both prescription and over-the-counter drugs, employee absentee rates, and emergency department visit data are all examples of electronic data collected for operational purposes that have been used successfully for syndromic surveillance.<sup>5-7</sup> For example, pharmacy sales can indicate the onset of community influenza activity before it appears in laboratory data.<sup>7</sup> The secondary use of health data is an active area of public policy discussion, <sup>1-2, 8-9</sup> particularly in light of the federal electronic health record (EHR) incentive program designed to increase adoption of EHRs.<sup>10</sup> For example, clinical data could assist in identifying patients eligible for pharmaceutical clinical trials, providing a potential revenue source for the sustainability of EHRs.<sup>11</sup>

Nevertheless, data collected for one purpose are rarely ideally suited for secondary use. Data are frequently of variable quality, and missing data may be common. Manual processing may be needed to assess quality and standardize data formats for analysis.<sup>6</sup> Lack of data standards, or inconsistent application of them, may make it difficult or impossible to analyze data without advanced natural language processing techniques.<sup>34</sup>

As part of a series of research and quality improvement projects, we began examining data from a project management system being used to track the progress of electronic health record (EHR) implementations by a regional extension center. This system, hosted by Salesforce (Salesforce.com, Inc., San Francisco, CA), contains information about several thousand clinicians and practices that are receiving EHR implementation support from the Primary Care Information Project (PCIP) at the New York City Department of Health and Mental Hygiene.<sup>12</sup>

As we examined the project management data, we encountered a variety of data quality issues reminiscent of larger issues in secondary use. In this paper, we identify and describe these challenges, trace them to their root causes, and place them in context of similar issues in the literature on secondary use.

## Background

The Primary Care Information Project is an initiative of the New York City Department of Health and Mental Hygiene with the mission of improving the delivery of health care in ambulatory settings through promoting adoption and use of EHRs in New York City. PCIP purchases EHR software licenses on behalf of eligible providers, subsidizes maintenance and support costs for 2 years, manages implementation processes in cooperation with the EHR vendor, and provides additional post-go-live EHR training and support with a focus on quality improvement. In 2010, the Fund for Public Health in New York won a federal regional extension center (REC) award and established the Regional Electronic Adoption Center for Health (REACH), a program under PCIP.

Since early 2007, a web-based project management system product by Salesforce.com has been used to track implementations. The project management system is used routinely by multiple PCIP teams. For example, outreach staff collect information about clinicians potentially interested in implementing an EHR, and document ongoing contacts with them. In addition, members of the implementation team use the system as they launch the EHR implementation for each small practice, collect additional descriptive information, and document key project milestones. These implementation staff capture a variety of descriptive information about each practice in structured and free-text data fields, attach documents to the record, and use free-text fields to write notes about telephone calls, questions, unresolved problems, and to-dos. Some of the many milestones recorded in the database in structured format include the date the contract was signed, the date of the so-called "kickoff call" at which the project plan was agreed upon, the dates of EHR and practice management system training sessions, and the EHR go-live date. After the EHR implementation process, a team of quality improvement staff use the same database to document training and assistance provided to clinicians and office staff.

As a result, the database contains descriptive records for individual people, as well as a complex set of longitudinal records for healthcare organizations. Currently, the database contains information about more than 2500 healthcare providers, 600 small private physician practices, and 30-plus community health centers, as well as 4 hospital outpatient departments at various stages of EHR implementation.

### Methods

We began examining the project management database for several purposes. First, we were interested in studying the challenges associated with EHR implementation among PCIP's participating practices and providers (an ongoing study being reported elsewhere). In addition, we had quality improvement goals for improving PCIP project management procedures.

For the ongoing EHR implementation study described above, we identified more than 30 variables in Salesforce with the potential to be relevant to the outcomes under evaluation. We used the Salesforce.com report tools to query the database for these variables, and computed frequencies to determine the rates of missing data. In cases when different variables had an obvious relationship to each other, we computed crosstab frequencies in order to identify inconsistencies and potential errors, such as a situation in which number of provider full-time equivalents (FTEs) was greater than the number of healthcare providers, or when the start date for a project was recorded as occurring after its end date. In addition, during all analyses, we tracked occurrence of any duplicate records (practices occurring in the database more than once). We held a series of weekly team meetings over about 4 months with key informants involved in the data collection to trace the root causes of these data quality issues and, in some cases, to develop data remediation plans.

The current study was part of a larger study of EHR implementation at PCIP being conducted as part of HITEC (the Health Information Technology Evaluation Collaborative), an academic consortium designated by the state of New York as the evaluation entity for health IT projects funded under the Health Care Efficiency and Affordability Law for New Yorkers capital grants program. The study was approved by the Weill Cornell Medical College Institutional Review Board.

## Results

We present 4 illustrative data quality issues and their root causes from one data set of small community practices participating in EHR implementations. These data quality issues were selected for presentation because resolving them was critical before the data could be used for secondary purposes, and because they appeared to illustrate more generalizable issues.

**DATA QUALITY ISSUE 1:** In this data set, 544 small practices had signed a contract to join PCIP. Of these, 430 (79%) had a recorded EHR go-live date, indicating that they had completed their EHR implementation; the remaining 114 (21%) were still in the process of implementation. However, 265 of the 430 were either missing the date upon which implementation started ("kickoff call"), or had inconsistent dates in different data fields of the database.

**Primary and secondary uses of these data:** The primary use for which these dates were collected was to establish a project plan for a practice, then document its progress. The order of milestone dates was more important than specific times between them; for example, the contract had to be signed before any of the subsequent milestones. However, for secondary use, the dates became important as markers for duration of implementation and its components.

**Root causes of data quality problems:** In tracing the data quality issue to the root causes, it became clear that different dates had different interested stakeholders, as well as different financial and contractual implications. Specifically, all the project stakeholders needed access to the correct contract signed date because it marked the start of the small practice's two-year software license. As a result, this date as recorded in the database was highly reliable.

By contrast, the kickoff call was originally a process that launched a series of events, and only later was identified as an operational start point that marked the begin date of implementation. As a result, as part of PCIP process improvement, PCIP worked with the EHR vendor to retrospectively capture the kickoff call date in records where it had not originally been captured. This required the EHR vendor staff to double-enter data into their own project management system and into the Salesforce system, leading to the potential for inconsistent data. During this retrospective data entry process, documentation practices varied, with some of the staff documenting the actual event date and others documenting the originally planned kickoff date, which was not always corrected if the kickoff call date was rescheduled.

In addition, the Salesforce database was constructed in such a way that there were two fields in that reflected the kickoff call date (the date field, and a "stage history" field). Although this flexibility was meant to provide better documentation capabilities for the users, it led to inconsistencies because the fields were not linked, and neither was definitively identified as the gold standard.

**Remediation plan:** Several members of the implementation team manually reviewed the dates, supporting documents, and free text notes in each practice's electronic record to determine when the "kickoff call" had actually occurred. The resulting gold standard list was subsequently used to correct the data in the database.

**DATA QUALITY ISSUE 2:** Of the 544 small practices, 31 (5.7%) were documented to have had a previous EHR before joining PCIP, 236 (43.4%) indicated they did not have an EHR, and the remaining 277 (50.9%) had missing data.

**Primary and secondary uses of these data:** These data were collected as part of an application form that assessed the practice's eligibility for the PCIP program as well as its perceived readiness for the new technology. The perceived readiness questions included questions about previous exposure to EHRs and other technologies. The secondary use of these data was as an indicator of a practice's experience with technology, which might correlate with the speed or ease of the implementation.

*Root causes of data quality problems:* The missing data problem originated in several changes in the department's procedures pertaining to the recruitment of new practices, which led to corresponding configuration changes in the electronic systems.

In the early years of the EHR implementation program, physician practices completed a paper application form, which was sometimes input into the database by project manager but other times was scanned and attached as a PDF to the project management record, where it could be consulted by any PCIP staff member. However, later, the department implemented an online application form linked directly to the Salesforce database, so that the questionnaire answers automatically populated the database. As a result, our initial attempt to export the questionnaire answers for analysis revealed large quantities of missing data in the structured fields.

An additional challenge was that in several cases, the PDF was not linked to the practice's electronic project management record but rather to the electronic record for the practice employee who had completed the questionnaire. This was most likely because at the time the application was submitted, an electronic record had not yet been created for the practice.

*Remediation plan:* A student intern retrieved the PDFs where available and manually input the questionnaire data into the appropriate fields of the database.

**DATA QUALITY ISSUE 3**: In our initial data query, we identified several practices with the same name but different PCIP-assigned ID numbers, as well instance in which as the same ID number was assigned to practices with different names.

*Primary and secondary uses of these data:* The PCIP ID number was originally assigned to track each practice in terms of their service contract and linked this contract to their name as originally entered. For secondary use, the PCIP ID number became the way that all entries about any practice were linked for tracking and trending over time.

**Root causes of data quality issues:** Most PCIP staff tended to use the practice name as its identifier, rather than the PCIP ID number. Although this did result in the ad hoc development of a standardized vocabulary of practice names, practice names still occasionally varied, especially for newly enrolled practices. Over time, some practices changed their names, merged, split, or closed, leading to duplicate records. Duplicate PCIP ID numbers occurred in a very small number of cases, most of which were traced to preliminary contacts with practices that did not end up enrolling with PCIP and that were associated with an almost entirely empty electronic record.

Remediation plan: Manual review of records successfully disambiguated all of the cases.

**DATA QUALITY ISSUE 4**: A database field entitled "number of providers" for a practice yielded a different number than was produced by a count of individual provider records linked to the practice record.

*Primary and secondary uses of these data:* For primary use, the number of providers was helpful in developing the project plan as well as tracking completion of milestones such as provider training. For secondary use, the number of providers was collected as a potential predictor of implementation time for the entire EHR implementation.

*Root causes*: The "number of providers" field originated from the value on the application questionnaire, which was either self-reported by the practice or estimated by a PCIP outreach team member. At best, it represented an estimate of practice's staffing level before joining PCIP for very rough planning purposes. However, the electronic database records associated with the individual providers reflected the actual number of EHR software licenses issued upon joining PCIP. This number was determined to be more reliable, as external stakeholders (in this case the software vendor) needed to know the number of software licenses.

*Remediation plan:* No remedial actions were taken, but we determined to use the provider count for future analyses rather than the "number of providers" field.

## Discussion

A large project management data system used in a consistent fashion for 4 years provided a rich data set for secondary uses including research and quality improvement. Nevertheless, early experiences using this data for secondary purposes revealed considerable variability in quality and integrity of the data. In our exploration of root causes, we determined that these variations in data quality arose from:

- **Differential incentives for the accuracy of the data**. Data were documented consistently if they had had financial or contractual implications and were of interest to external stakeholders such as lawyers, the software vendor, or the clinician clients of PCIP, whereas data being used solely for internal purposes showed more variability.
- *Flexibility in system software that allowed multiple routes to documenting the same tasks.* For example, two structured fields were available for documenting a particular milestone date, and the application questionnaire was accepted as both PDF and structured data.
- *Variability in documentation practices among different personnel documenting the same task.* For example, a particular date field could be used to document either the scheduled date of an event or the actual date of that event.
- *Variability in use of standardized vocabulary*, specifically, the internally developed standardized vocabulary of practice names.
- *Changes in project procedures and electronic system configuration over time*, as when a paper questionnaire was replaced with an electronic version.

A larger issue linking all of these observations was that our secondary use of data, which required aggregating historical data within each practice and also across practices, required a different and generally higher degree of data integrity than was required for the original primary use. Staff members could successfully manage EHR implementation even with imperfect database data because this database was only one source of information: project managers were also immersed in a rich ongoing stream of information from meetings, telephone calls, e-mail, site visits, and paper documents. In addition, the sequential nature of project management meant that pieces of data in the database. Finally, in this decentralized system, a single project manager took responsibility for a single practice throughout the implementation process. Idiosyncratic ways of entering data thus had no serious impact, as a single person was both the source of data input and the audience for that data.

Although the current data set included no health data, the issues we have identified map closely to previously identified problems in clinical data quality that pose challenges for secondary uses such as research and quality reporting. Botsis and colleagues<sup>3</sup> have identified such issues in particularly granular detail in a description of their use of the Columbia University Medical Center clinical data warehouse to conduct a survival analysis of pancreatic cancer patients. Although they were able to use database queries to retrieve information, they also had to do significant manual review and data abstraction, including manual review of free-text notes to ensure the accuracy of the extracted data. Incomplete, inconsistent, and inaccurate data were common; in some patient subsets, important variables had more than 50% missing values. The authors did not do a formal root cause analysis, but were able to identify potential causes. For example, inconsistencies arose when the same data were being entered into different fields of a single EHR,<sup>3</sup> just as we observed in cases when the Salesforce database offered multiple alternatives for documenting the same information. Botsis et al also noted that missing and inconsistent data were common. This may have been because the information needed to document treatment may not have included the types of disease progression events that were of interest from a secondary use perspective. Dates were particularly likely to be missing; the difficulty of accurately interpreting temporal information in clinical data is a well-known problem.<sup>13</sup>

The issues recorded here, their root causes, and potential solutions were not evident from inspection of the database. Rather, they emerged only after intensive and collaborative discussions among researchers and the project managers with primary responsibility for data entry. Explanations for the data quality issues and novel ways of analyzing the data emerged only from in-depth understanding of the daily workflow being documented in the project management system and the history of the organization.

We conclude that researchers interested in secondary use of data must immerse themselves in the workflow processes being documented in order to understand the data and reasons for problems. In addition, organizations that may be interested in secondary uses of data will benefit from close attention to documentation practices, including

incentivizing the documentation of important tasks, eliminating redundancy in data fields, ensuring consistent data definitions, and promoting uniform standards and training for those involved in documentation. As others have noted, "no purely technical solution can overcome the capture of inaccurate information by the user of a clinical information system. As such, nontechnical innovations that help improve the accuracy of recorded information and incentivize consistently accurate data collection are critical to the success of research initiatives that rely on the presence of such data."<sup>1</sup>

## Acknowledgments

We are grateful to Maryam Kahn for data entry, and to Anupam Kashyap for consultation on some of the root causes of the data quality issues. This study was supported by the New York State Department of Health (NYS contract number C023699).

## References

- 1. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement. *Annals of Internal Medicine*. 2009;151(5):359-360.
- 2. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: An American Medical Informatics Association white paper. *JAMIA*. 2007;14(1):1-9.
- **3.** Botsis T, Hartvigsen G, Chen F, Weng C. Secondary uses of EHR: Data quality issues and informatics opportunities. *AMIA Joint Summits on Translational Science Proceedings*. 2010;2010.
- 4. Prokosch H-U, Ganslandt T. Reusing the electronic medical record for clinical research. *Methods of Information in Medicine*. 2009;48:38-44.
- 5. Heffernan R, Mostashari F, Das D, et al. System descriptions: New York City syndromic surveillance systems. *Morbidity and Mortality Weekly Report*. 2004;53 (Suppl):23-27.
- 6. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. *Emerging Infectious Diseases*. 2004;10(5):858-864.
- 7. Increased antiviral medication sales before the 2005-06 influenza season New York City. *MMWR Morbidity and Mortality Weekly Report*. 2006;55:277-279.
- **8.** Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel system a national resource for evidence development. *New England Journal of Medicine*. 2011;364(6):498-499.
- **9.** President's Council of Advisors on Science and Technology. *Report to the president Realizing the full potential of health information technology to improve health care for Americans: the path forward.* Washington, DC: Executive Office of the President of the United States,;2010.
- **10.** Medicare and Medicaid Programs. *Electronic Health Record Incentive Program, Final Rule*: 75 Federal Register 144 (28 July 2010);2010.
- **11.** Miller JL. The EHR solution to clinical trial recruitment in physician groups. *Health Management Technology*. 2009;<u>http://www.healthmgttech.com/</u>.
- 12. Mostashari F, Tripathi M, Kendall M. A tale of two large community electronic health record extension projects. *Health Aff (Millwood)*. Mar-Apr 2009;28(2):345-356.
- **13.** Hripcsak G, Elhadad N, Chen Y, Zhou L, Morrison F. Using empiric semantic correlation to interpret temporal assertions in clinical texts. *Journal of the American Medical Informatics Association* 2009;16(To):220-227.

# Developing the Sentinel System — A National Resource for Evidence Development

Rachel E. Behrman, M.D., M.P.H., Joshua S. Benner, Pharm.D., Sc.D., Jeffrey S. Brown, Ph.D., Mark McClellan, M.D., Ph.D., Janet Woodcock, M.D., and Richard Platt, M.D.

he Food and Drug Administration (FDA) now has the capacity to "query" the electronic health information of more than 60 million people, posing specific questions in order to monitor the safety of approved medical products. This pilot program, called Mini-Sentinel, uses a distributed data network (rather than a centralized database) that allows participating health plans and other organizations to create data files in a standard format and to maintain possession of those files. These organizations perform most analyses of their own data by running computer programs distributed by a coordinating center, and they provide consistent summarized results for the FDA's review.1 The principles and practices involved in this effort to improve the safety of medical products can inform other uses of electronic health information to answer additional important questions about health and health care.

When the FDA announced the Sentinel Initiative in May 2008, it established a vision and objectives for the program, including the development of the Sentinel System, which will eventually be able to search the electronic health data of a minimum of 100 million patients.<sup>2</sup> Laying the groundwork for that system has required an extraordinary range of input from public and private organizations. Under a cooperative agreement with the FDA, the Engelberg Center for Health Care Reform at the Brookings Institution has been convening an ongoing series of discussions among stakeholders to address the near- and long-term

challenges inherent in implementing the Sentinel System.3 In 2009, the FDA gave the Harvard Pilgrim Health Care Institute the lead role in fulfilling a 5-year contract to establish a system - the Mini-Sentinel — for developing and testing approaches and methods that could be used to inform the structure and operations of the full Sentinel System. The institute is now leading a diverse partnership of approximately 200 epidemiologists, clinical content experts, statisticians, and data specialists from 27 institutions that are participating in this pilot system (www.minisentinel.org).

Through the Mini-Sentinel, capabilities are being developed for actively monitoring the safety of approved medical products using the electronic health information in claims systems, inpatient and outpatient medical records, and patient registries. The Mini-Sentinel builds on the work of the Vaccine Safety Datalink project (managed by the Centers for Disease Control and Prevention), the HMO Research Network, the Population Medicine Distributed Research Network (PopMedNet, funded by the Agency for Healthcare Research and Quality), and the Observational Medical Outcomes Partnership, among others.4

In the first year of the Mini-Sentinel project, its leaders established a network of data partners and a system with robust patientprivacy policies that could be used in querying the network's databases. The initiative's distributed data network allows each data partner to maintain physical and operational control over its own

patient-level data, while providing the aggregated information needed to address the FDA's questions. Source data reside behind the data partners' institutional firewalls, where they are transformed into a standard format. This approach allows each data partner to answer the FDA's queries by executing standardized computer programs distributed by the Mini-Sentinel Operations Center. A typical result might include the number of new users of a product who experience a particular outcome, grouped according to age, sex, other treatments, and health status. This use of distributed analysis - whenever possible - eliminates or greatly reduces the exchange of protected health information. The data partners can obtain full-text medical records when necessary to confirm diagnoses or exposures and to determine the existence or severity of risk factors.

The initial focus of Mini-Sentinel has been on developing the ability to use claims data. In the next year, laboratory-test results and vital signs, derived from electronic health records and clinical laboratory records, will be added. The partnership is also evaluating procedures whereby Mini-Sentinel data partners will be able to link to data held by other organizations, such as state immunization registries and device registries.

The FDA will soon begin to actively monitor the data, seeking answers to specific questions about the performance of medical products, such as the frequency of myocardial infarction among users of oral hypoglycemic agents (a topic selected because it has

N ENGLJ MED 364;6 NEJM.ORG FEBRUARY 10, 2011

The New England Journal of Medicine

Downloaded from nejm.org by RICHARD PLATT on September 25, 2011. For personal use only. No other uses without permission.

Copyright © 2011 Massachusetts Medical Society. All rights reserved.

been difficult to identify druginduced myocardial infarction through existing prospective surveillance mechanisms). The FDA will also monitor the occurrence of adverse events associated with select routinely administered vaccines. Using the Mini-Sentinel system, the FDA will also be able to obtain rapid responses to new questions about medical products and, eventually, to evaluate the health effects of its regulatory actions. This monitoring portfolio will expand as the FDA and its collaborators acquire experience and develop operational efficiencies and as additional data resources become available.

The distributed-database-andanalysis model and the infrastructure of the Mini-Sentinel data network can be extended to other forms of evidence development. Provisions in the economic stimulus and health care reform legislation, and a recent report from the President's Council of Advisors on Science and Technology,5 envision expanded use of electronic health information for other types of public health surveillance, quality measurement, comparative effectiveness research, and biomedical research — all of which are essential to improving the country's health and health care delivery system.

Issues relevant to other secondary uses of electronic health information include recruitment of appropriate data partners, development and refinement of analytic methods, implementation of standards to ensure that analytic methods are consistent across the data sources, and above all, protection for the rights and privacy of patients. Data privacy and security are top priorities that were key considerations in the decision to build Mini-Sentinel as a system that uses a distributed data system and distributed analysis whenever possible. The committed collaboration among representatives of patients and consumers, health care professionals, Mini-Sentinel's data partners and safety scientists, and the medical-products industry has been essential to the Sentinel Initiative's progress.

It is particularly challenging to establish appropriate governance for a distributed data network that can support multiple secondary uses for health information. The current infrastructure is supported by a single federal agency, the FDA, and all the data are provided by private organizations, yet potential users of such a system reside not only broadly in government but also in academia, the private sector, and other user communities. To facilitate the development of this infrastructure into a national resource, this distributed system may ultimately be best managed by a consortium of interested parties operating as a public-private partnership. For example, specialized network-coordinating centers might rely on a consistent infrastructure to use the same sources of health information for various purposes, including public health uses, effectiveness research, quality measurement, and health services research.

The envisioned Sentinel System will build on the knowledge, partnerships, data resources, privacy protections, and technical capabilities that are being developed in the Mini-Sentinel program. Success in the form of improved safety of medical products will depend on the continued engagement of all concerned stakeholders and on ensuring that patients, consumers, and health care providers understand that all medical products pose risks and that postmarketing surveillance is critical to expanding the limited evidence base that exists when products are approved. Success also depends on the continued development of surveillance methods and on increasing the workforce of scientists who are trained to develop and interpret this evidence effectively.

Health care data represent a precious resource that must be used to the fullest possible extent to promote the public health, while the rights of patients and consumers are protected. As an early working model for secondary uses of data produced in the routine delivery of health care, the Sentinel System can and should become a national resource for evidence development and a cornerstone of a learning health care system.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Food and Drug Administration, Silver Spring, MD (R.E.B., J.W.); the Engelberg Center for Health Care Reform, Brookings Institution, Washington, DC (J.S. Benner, M.M.); and the Department of Population Medicine, Harvard Pilgrim Health Care Institute, and Harvard Medical School both in Boston (J.S. Brown, R.P.).

This article (10.1056/NEJMp1014427) was published on January 12, 2011, at NEJM.org.

1. Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network — improving the evidence of medicalproduct safety. N Engl J Med 2009;361:645-7.

 The Sentinel Initiative: national strategy for monitoring medical product safety. Silver Spring, MD: Food and Drug Administration, May 2008. (http://www.fda.gov/downloads/ Safety/FDASSentinelInitiative/UCM124701.pdf.)
 Summaries of Brookings events on active medical product surveillance. Washington, DC: Brookings Institution. (http://www.brookings .edu/health/Projects/surveillance.aspx.)

4. 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:Suppl:S45-S51.

5. Realizing the full potential of health information technology to improve healthcare for Americans: the path forward. Washington, DC: Executive Office of the President, December 2010. (http://www.whitehouse.gov/sites/default/ files/microsites/ostp/pcast-health-it-report.pdf.) *Copyright* © 2011 Massachusetts Medical Society.

The New England Journal of Medicine

Downloaded from nejm.org by RICHARD PLATT on September 25, 2011. For personal use only. No other uses without permission.

Copyright © 2011 Massachusetts Medical Society. All rights reserved.

# Rapid Identification of Myocardial Infarction Risk Associated With Diabetes Medications Using Electronic Medical Records

John S. Brownstein, phd<sup>1,2,3</sup> Shawn N. Murphy, md, phd<sup>4,5</sup> Allison B. Goldfine, md<sup>5,6</sup> Richard W. Grant, md<sup>7,8</sup> Margarita Sordo, phd<sup>4,5</sup> Vivian Gainer, ms<sup>5</sup>

Judith A. Colecchi, ms<sup>5</sup> Anil Dubey, md<sup>4,5</sup> David M. Nathan, md<sup>8,9</sup> John P. Glaser, phd<sup>5</sup> Isaac S. Kohane, md, phd<sup>1,5,10</sup>

**OBJECTIVE** — To assess the ability to identify potential association(s) of diabetes medications with myocardial infarction using usual care clinical data obtained from the electronic medical record.

**RESEARCH DESIGN AND METHODS** — We defined a retrospective cohort of patients (n = 34,253) treated with a sulfonylurea, metformin, rosiglitazone, or pioglitazone in a single academic health care network. All patients were aged >18 years with at least one prescription for one of the medications between 1 January 2000 and 31 December 2006. The study outcome was acute myocardial infarction requiring hospitalization. We used a cumulative temporal approach to ascertain the calendar date for earliest identifiable risk associated with rosiglitazone compared with that for other therapies.

**RESULTS** — Sulfonylurea, metformin, rosiglitazone, or pioglitazone therapy was prescribed for 11,200, 12,490, 1,879, and 806 patients, respectively. A total of 1,343 myocardial infarctions were identified. After adjustment for potential myocardial infarction risk factors, the relative risk for myocardial infarction with rosiglitazone was 1.3 (95% CI 1.1–1.6) compared with sulfonylurea, 2.2 (1.6–3.1) compared with metformin, and 2.2 (1.5–3.4) compared with pioglitazone. Prospective surveillance using these data would have identified increased risk for myocardial infarction with rosiglitazone compared with metformin within 18 months of its introduction with a risk ratio of 2.1 (95% CI 1.2–3.8).

**CONCLUSIONS** — Our results are consistent with a relative adverse cardiovascular risk profile for rosiglitazone. Our use of usual care electronic data sources from a large hospital network represents an innovative approach to rapid safety signal detection that may enable more effective postmarketing drug surveillance.

Diabetes Care 33:526-531, 2010

From the <sup>1</sup>Children's Hospital Informatics Program at the Harvard–MIT Division of Health Sciences and Technology, Boston, Massachusetts; the <sup>2</sup>Division of Emergency Medicine, Children's Hospital Boston, Boston, Massachusetts; the <sup>3</sup>Department of Pediatrics, Harvard Medical School, Boston, Massachusetts; the <sup>4</sup>Laboratory of Computer Science, Massachusetts General Hospital, Boston, Massachusetts; the <sup>5</sup>Partners Healthcare, Boston, Massachusetts; the <sup>6</sup>Joslin Diabetes Center, Boston, Massachusetts; the <sup>7</sup>Division of General Medicine, Massachusetts General Hospital, Boston, Massachusetts; the <sup>8</sup>Department of Medicine, Harvard Medical School, Boston, Massachusetts; the <sup>9</sup>Diabetes Center, Massachusetts General Hospital, Boston, Massachusetts; and the <sup>10</sup>Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts.

Corresponding author: John S. Brownstein, john\_brownstein@harvard.edu.

Received 13 August 2009 and accepted 2 December 2009. Published ahead of print at http://care. diabetesjournals.org on 15 December 2009. DOI: 10.2337/dc09-1506.

J.S.B. and S.N.M. contributed equally to the work.

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

© 2010 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See http://creativecommons. org/licenses/by-nc-nd/3.0/ for details.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

dverse events that occur infrequently during premarketing randomized clinical trials or are underreported with traditional postmarketing methods of drug surveillance underscore the need for additional methodologies and data sources to monitor drug safety (1). Critical insights may be realized by monitoring large clinical databases using automated data feeds in near real time (2). Diabetes medications present an ideal paradigm to test new safety signal detection approaches because they are used frequently in large numbers of patients with type 2 diabetes, and new products have been recently launched while suitable drug comparators remain marketed. Existing concerns regarding adverse cardiovascular risk for diabetes therapies provide motivation for hypothesis-driven prospective surveillance. Adverse cardiovascular side effects have been seen with rosiglitazone (3,4). Although a recent noninferiority clinical trial has provided some evidence exonerating rosiglitazone from a risk for excess mortality (5), concern remains regarding a possible adverse risk for myocardial infarction.

We tested an automated strategy analyzing clinical data in real time to detect adverse drug-related events. Because premarketing clinical trials of diabetes therapies are currently designed primarily to evaluate efficacy for glycemic improvement and have not previously been designed to assess relatively infrequent but clinically important adverse outcomes, active surveillance may play a valuable role in assessment of risk (6). Active surveillance could provide evidence of risk earlier than postmarketing outcome trials. Furthermore, it may be cost prohibitive to conduct randomized controlled trials for each drug product toward important hard safety outcomes. Although such an analysis would not provide conclusive causal evidence, we determined whether prospective analysis of clinical data could have provided early evidence of cardiovascular risk associated with rosiglitazone that would warrant additional evaluation.

# **RESEARCH DESIGN AND**

**METHODS** — We identified a cohort of patients who had new prescriptions for diabetes medications within Partners Healthcare System, a large, nonprofit academic health care network including Brigham and Women's and Massachusetts General Hospitals. The source of clinical data was the Research Patient Data Registry, a centralized data warehouse including patient demographic information, dates of service, medications, diagnoses, laboratory results, and discharge summaries.

The retrospective cohort analysis included all patients aged >18 years identified by an ICD-9 code for Diabetes Mellitus (250.XX) or an A1C of >6.0%and at least one record of prescription of an oral diabetes medication as an outpatient or dispensation as an inpatient, between 1 January 2000 and 31 December 2006. Analyses focused on three classes of diabetic medications: sulfonvlureas, the biguanide metformin, and the thiazolidinediones, rosiglitazone and pioglitazone. Evidence of insulin therapy did not exclude patients but was adjusted for in multivariate models and used for stratified analysis (described below). We excluded patients receiving either metformin or thiazolidinedione who had a diagnosis of polycystic ovaries but not diabetes. For each patient, all available associated data were extracted, including narrative notes and hospital discharge summaries. Narrative notes were used for validating coded medications and diagnoses found in medical records, permitting determination of sensitivity and specificity of events as recorded in the electronic medical record.

### Patient enrollment, observation, drug exposure, and event identification

The study population does not receive health care exclusively within the Partners system, and, thus, some patients within the surveillance database may have had incomplete records. To address this issue, we used health care encounters (inpatient or outpatient) as a proxy for receipt of care at Partners over a specific observation period. We constructed 14 6-month observation periods, beginning on 1 January or 1 July between 2000 and 2006, during which a patient had at least one outpatient office visit, including psychotherapy or nutrition visits, or an inpatient encounter. Study entry was considered the first period meeting one of these criteria within the study dates.

For each patient, duration of exposure to individual diabetes medications was assessed in 6-month increments during which only one of the four medications was prescribed. Patients receiving multiple medications under consideration were excluded. The study end point for each evaluable patient was first hospitalization between 1 January 2000 and 31 December 2006 for myocardial infarction (ICD-9 code 410), death (all causes), a gap in care in which there were no patient encounters in subsequent observation periods, or end of study in 2006. The ICD-9 diagnostic code for acute myocardial infarction has been validated previously (7). Events were associated with a particular medication only when the prescription or dispensation occurred within the 6 months before the documented myocardial infarction. If a patient did not have any activity for a 6-month observation period but resumed activity in the following period, than the particular 6-month observation period with no activity was excluded from analysis. Analysis was repeated considering only patients having been prescribed one of the four medications, considered to be monotherapy. Finally, we also performed stratification of our data to analyze patients who had not received insulin as outpatient therapy.

We conducted a manual review of outpatient notes and inpatient discharge summaries on a random sample of 200 patients to validate use of electronic medical record data to identify both drug exposure and myocardial infarction events. Review included patients identified as exposed to rosiglitazone and with myocardial infarction (n = 50) or exposed and without an event (n = 50) as well as the comparator group of patients (receiving one of the other three oral diabetes medications but not exposed to rosiglitazone) and with (n = 50) or without myocardial infarction event (n = 50). Institutional review board approval was obtained for medical record review.

### Statistical analysis

The relative risk of myocardial infarction associated with therapy was calculated for rosiglitazone compared with metformin, sulfonylureas, or pioglitazone. Both crude and adjusted rate ratios with 95% CIs were estimated using generalized linear modeling, assuming a Poisson distribution for the response and set duration of time taking a particular medication (as 6-month intervals) as the offset. To account for overdispersion in the count data, extra-Poisson variability was modeled and incorporated into estimates of SEs. Parameter estimates were transformed to rate ratios.

Adjustments were made for potential risk factors including age, sex, cardiovascular disease prior to enrollment (defined by billing codes for coronary artery disease, myocardial infarction, angina, congestive heart failure, cerebrovascular incident, percutaneous coronary intervention, and coronary artery bypass graft surgery), any use of hypertensive medications, lipid-lowering medications, and outpatient insulin use during study period. The model also included adjustment for underlying morbidity using an ageadjusted Charlson score. In an additional model, we evaluated potentially important factors for which we had less than complete data. These included race/ ethnicity (with information available in 93% of patients), insurance coverage (commercial, Medicare, Medicaid, or uninsured) (83%), A1C (60%), and creatinine (71%) levels. Overall mean A1C and creatinine levels (<2.0 or  $\geq$ 2.0 mg/dl) during the study period were considered indicators of diabetes severity. Differences in these characteristics between medication groups were identified with ANOVA and a Tukey post hoc test. Finally, because previous myocardial infarction imparts a greater risk for recurrent cardiovascular events (8) and because of the need to consider starting new medications to minimize potential prolonged effects of prior diabetes therapies on cardiovascular events, we tested a model in which all patients who had ever had a recorded inpatient stay for myocardial infarction or had been prescribed a diabetes medication in the year before entry were excluded.

## Signal detection analysis

To construct a general surveillance approach to identify adverse events from clinical data, we repeated the above analysis using a cumulative temporal approach by the defined 6-month intervals. All available data from the first time period (1 January 2000-31 May 2000) were analyzed, and data were iteratively added with each subsequent 6-month period. Cumulative data were analyzed until the final period. Data were treated as cumulative with additional new patients and patient-year exposure providing increased power to the analyses. A significant risk ratio (where the lower bound of the 95% CI was >1.0) was considered to

## Table 1—Characteristics of the population

	Rosiglitazone	Metformin	Sulfonylurea	Pioglitazone
n	1,879	12,490	11,200	806
Age*	$64.0 \pm 11.4$	$61.7 \pm 12.2$	$65.8 \pm 12.1$	$63.7 \pm 11.5$
Female sex	908 (48.3)	6,628 (53.1)	4,760 (42.5)	384 (47.6)
Myocardial infarction outcome	133 (7.1)	406 (3.3)	768 (6.9)	36 (4.5)
Prior myocardial infarction	234 (12.5)	1,421 (11.4)	1,945 (17.4)	94 (11.7)
Prior cardiovascular disease	597 (31.8)	3,369 (27.0)	4,544 (40.6)	251 (31.1)
Hypertension	1,689 (89.9)	10,454 (83.7)	10,076 (90.0)	709 (88.0)
Hyperlipidemia	1,466 (78.0)	8,484 (67.9)	7,545 (67.4)	602 (74.7)
Chronic renal insufficiency (creatinine >2 mg/dl)	338 (18.0)	936 (7.5)	2,374 (21.2)	121 (15.0)
Outpatient insulin use	446 (23.7)	2,341 (18.7)	1,425 (12.7)	263 (32.6)
AIC	$8.0 \pm 1.7$	$7.8 \pm 1.7$	$7.7 \pm 1.7$	$8.1 \pm 1.8$
Antihyperlipidemic medication use	1,340 (71.3)	7,721 (61.8)	6,610 (59.0)	556 (69.0)
Combination	35 (1.9)	90 (0.7)	80 (0.7)	19 (2.4)
Fibrates	191 (10.2)	887 (7.1)	730 (6.5)	86 (10.7)
Statins	1,287 (68.5)	7,473 (59.8)	6,428 (57.4)	526 (65.3)
Antihypertensive medication use	1,535 (81.7)	9,358 (74.9)	8,620 (77.0)	649 (80.5)
ACE inhibitors	1,096 (58.3)	7,019 (56.2)	6,108 (54.5)	463 (57.4)
Angiotensin II antagonists	406 (21.6)	1,931 (15.5)	1,697 (15.2)	170 (21.1)
β-Blockers	1,033 (55.0)	5,490 (44.0)	6,138 (54.8)	400 (49.6)
Calcium channel blockers	551 (29.3)	2,783 (22.3)	3,219 (28.7)	204 (25.3)
Combinations	275 (14.6)	1,844 (14.8)	1,315 (11.7)	123 (15.3)
α-β	163 (8.7)	739 (5.9)	1,068 (9.5)	47 (5.8)
Potassium-sparing diuretics	5 (0.3)	67 (0.5)	73 (0.7)	4 (0.5)
Unclassified combinations	10 (0.5)	24 (0.2)	17 (0.2)	4 (0.5)
Age-adjusted Charlson score*	$7.9 \pm 4.4$	$7.1 \pm 4.2$	$8.5 \pm 4.5$	$7.5 \pm 4.2$

Data are n (%) or means  $\pm$  SD. \*Age at index date.

be a safety signal. All analyses were performed using SAS statistical software (version 9.0; SAS Institute, Cary, NC). Numbers of prescriptions of pioglitazone were insufficient for comparison with rosiglitazone until 1 January 2002.

**RESULTS** — We identified 34,252 diabetic patients treated with at least one of the four diabetes medications between 1 January 2000 and 31 December 2006. Of the total 159,586 evaluable 6-month intervals, there were 40,695 periods of sulfonylurea therapy (17,157 patients), 48,713 periods of metformin therapy (18,162 patients), 8,707 periods of rosiglitazone therapy (4,274 patients), and 3,591 periods of pioglitazone therapy (1,800 patients). When only one of the four diabetes medications was prescribed in a 6-month period, we identified 20,233 periods for sulfonylureas (11,200 patients), 27,860 periods for metformin (12,490 patients), 2,834 periods for rosiglitazone (1,879 patients), and 1,290 periods for pioglitazone (806 patients) (Table 1). When only one of the four diabetes medications was prescribed during the entire period, we identified 7,152 patients taking sulfonylureas, 8,798 patients

taking metformin, 1,028 patients taking rosiglitazone, and 418 patients taking pioglitazone. Given the large number of patients in the different treatment groups, there were statistically significant, although generally small, differences in many baseline variables. These were adjusted for in analyses to control for known baseline differences.

We identified 1,343 hospitalized myocardial infarction events and an overall event rate of 16.8 per 1,000 patientyears. There were 768 events associated with sulfonylureas (38.0 events per 1,000 patient-years), 406 with metformin (14.6 events per 1,000 patient-years), 133 with rosiglitazone (46.9 events per 1,000 patient-years), and 36 with pioglitazone (27.9 events per 1,000 patient-years). Manual review of 235 randomly selected patient records revealed a high level of confirmation for drug exposure to individual medications, with both sensitivity and specificity of 94%. Identification of myocardial infarction events was confirmed with a sensitivity of 93% and specificity of 74%. Lower specificity was primarily due to the presence of previous and "rule out" myocardial infarctions noted in patient records. Overall, there

were no differences in specificity and sensitivity of myocardial infarction by drug type.

Rosiglitazone was associated with an unadjusted rate ratio for increased myocardial infarction of 1.2 (95% CI 1.0–1.3) compared with sulfonylureas, 3.3 (2.9-3.6) compared with metformin, and 1.7 (1.3-2.1) compared with pioglitazone. After adjustment for identified risk factors (age, sex, cardiovascular disease, hypertensive medications, lipid-lowering medications, and age-adjusted Charlson score), individuals treated with rosiglitazone had an increased rate ratio for myocardial infarction risk of 1.3 (1.0-1.6) compared with sulfonylurea, 2.7 (2.2-3.4) compared with metformin, and 1.7 (1.1-2.6) compared with pioglitazone. Additional adjustments for factors with limited data in our patient population (race/ethnicity, insurance coverage, A1C, and creatinine levels) resulted in only small differences in adjusted relative risk. In the model with additional factors not available for the entire population, rosiglitazone was associated with a relative risk of myocardial infarction compared with sulfonylurea, metformin, and pioglitazone of 1.4 (95% CI 1.0-1.9), 2.4 (1.04.2), and 2.0 (1.0-4.2), respectively. Analyses restricted to patients without prior myocardial infarction (29,055 of 34,252) and patients with no prior diabetes medication in the 12 months before enrollment (30,142 of 34,252) had no effect on model results.

Considering only patients receiving monotherapy, rosiglitazone was associated with an unadjusted rate ratio for increased myocardial infarction of 1.1 (95% CI 1.0-1.3) compared with sulfonylureas, 3.5 (3.1-3.9) compared with metformin, and 1.9(1.4-2.5) compared with pioglitazone. After adjustment for identified risk factors, individuals treated with rosiglitazone had an increased rate ratio for myocardial infarction of 1.2(1.0-1.4)compared with sulfonylurea, 2.5 (2.0-3.2) compared with metformin, and 1.7 (1.3-2.2) compared with pioglitazone. In the model with additional factors not available for the entire population, rosiglitazone was associated with a relative risk of myocardial infarction compared with sulfonylurea, metformin, and pioglitazone of 1.3 (95% CI 1.1-1.6), 2.2 (1.6-3.1), and 2.2 (1.5-3.4), respectively.

After performing stratification of our data to analyze patients who had not received insulin as an outpatient therapy, we found that rosiglitazone was associated with an unadjusted rate ratio for increased myocardial infarction of 1.3 (95%) CI 1.1–1.4) compared with sulfonylurea and 3.5 (3.2-3.9) compared with metformin. After adjustment for identified risk factors, individuals treated with rosiglitazone had an increased rate ratio for myocardial infarction risk of 1.3 (1.0-1.7) compared with sulfonylureas and 3.0 (2.4-3.7) compared with metformin. In the model with additional factors not available for the entire population, rosiglitazone was associated with a relative risk of myocardial infarction compared with sulfonylureas and metformin of 1.4 (95% CI 1.0-2.0) and 2.6 (1.8-3.6), respectively. No myocardial infarctions were identified among the 594 patients receiving pioglitazone without additional insulin outpatient therapy.

The iterative temporal analysis to define the earliest possible date a safety signal would have been detected (Fig. 1) demonstrates that a safety signal would have been identified for rosiglitazone compared with metformin after 18 months in July 2001 with an adjusted risk ratio of 2.1 (95% CI 1.2–3.8). Compared with sulfonylurea or pioglitazone, rosiglitazone safety signals would have been



**Figure 1**—Temporal analysis to ascertain the calendar date for earliest identifiable risk associated with rosiglitazone compared with other therapies is shown with each curve representing relative risk ratio of myocardial infarction for patients on rosiglitazone compared with alternatively prescribed medications (sulfonylurea, metformin, and pioglitazone).

identified by January 2005 with adjusted risk ratios of 1.2 (1.1–1.8) and 1.8 (1.0–3.4), respectively.

**CONCLUSIONS** — A recent metaanalysis of available case-control and cohort studies derived from the rosiglitazone phase III clinical dataset suggested a 43% increased risk for cardiovascular events in patients receiving rosiglitazone (3). Many factors contribute to uncertainty regarding these findings, including availability of only summary trial-level data rather than patient-level data, heterogeneity of trial design, and absence of uniform event adjudication (9). However, review of patient-level data by the U.S. Food and Drug Administration (FDA) yielded similar relative risk findings (10). Absolute risk was low because cardiovascular event rates were sparse in these studies and statistical methods to deal with infrequent event rates yield uncertainty regarding validity of the risk (11). Likewise, phase IV studies in patients with type 2 diabetes have neither confirmed nor excluded an increased hazard ratio for rosiglitazone (12,13), and, similarly, large randomized multicenter trials in high-risk diabetic patients with substantial use of rosiglitazone neither confirm nor exclude increased risk (14,15). The recently completed phase IV Rosiglitazone Evaluated for Cardiovascular Outcomes Regulation of Glycaemia in Diabetes (RECORD) study was designed as a noninferiority study comparing rosiglitazone plus either sulfonylurea or metformin versus metformin and sulfonylurea. Although it was underpowered and treatment crossover complicated interpretation of findings, relative risk for mortality was  $\sim 1.0$ ; however, risk for myocardial infarction with rosiglitazone was 1.14, leaving the risk of rosiglitazone for myocardial infarction uncertain (5). In contrast, results of randomized phase IV clinical trials and meta-analyses have suggested pioglitazone to be neutral to favorable in cardiovascular risk profile (16,17).

The thiazolidinediones rosiglitazone and pioglitazone both gained FDA approval within a short time span, have similar indications for being prescribed, have similar cost, and are initially without apparent prescription bias. A comparison of these two products reduces the likelihood of comorbidities and unmeasured variables confounding findings, which might cause greater potential bias for drugs of different class, cost, or safety profiles. Thus, evaluating cardiovascular safety of approved oral diabetes therapies in a real-world setting provides context, internal model validation, and potentially valuable clinical information for health care providers.

Our results are consistent with a previously suggested protective effect for

## Myocardial infarction and diabetes medications

metformin (18), more neutral effect for pioglitazone (16,17), and potential relative adverse cardiovascular safety profile for rosiglitazone (19,3,4,20). In particular, our results comparing rosiglitazone with pioglitazone complement other recent findings (19,20) and are not likely to be confounded by indication, given the similar prescribing patterns. Together, these findings demonstrate that methods for medical record surveillance may provide useful adjunct methods to assess postmarketing drug safety.

It is interesting that the relative risk confidence limits boundary either touches or is near 1.0 in all analysis for rosiglitazone compared with sulfonylureas. Sulfonylureas are established agents that have been used for the treatment of type 2 diabetes since approval in the 1950s. Our findings are consistent with those in the meta-analysis performed by the FDA (10), which suggested no increase in risk for rosiglitazone compared with this established therapeutic class.

Our results do differ somewhat from other recent studies. An analysis of potentially more robust data showed similar trends of decreased relative risk for cardiovascular disease with metformin and increased relative risk for sulfonylureas, but they did not show significant increased relative risk for rosiglitazone compared with pioglitazone (21). However, this finding contrasts with other observational studies showing increased risk with rosiglitazone (19,22). Differences among patient populations, in absolute event rates, or in methodologies may underlie differences in the magnitude of relative risks in such studies.

Importantly, combined treatments for dyslipidemia, hypertension, antithrombotic agents, and glycemia have markedly reduced event rates in patients with type 2 diabetes, and these gains are realized using strategies that include rosiglitazone (13,14). Relative risk analysis may be used to inform a provider regarding priority for selecting among treatment options, but individual patient comorbidities and tolerance must also be considered when one is choosing among specific therapeutic options, and absolute risk must be carefully considered before withholding a therapeutic option.

Our analysis does have important limitations. We do not have complete longitudinal prescription data for all individual patients, and patients may not take medication that has been prescribed. Hence we cannot confirm for all patients

whether they were taking a medication at the time of myocardial infarction. Although we have derived an estimate of recent exposure, defining true exposure is currently not possible with usual clinical data. We may also have missed patients who did have exposure. Prescriptions for diabetes medications may have been obtained outside the Partners system and may therefore not have been captured. However, this situation would underestimate rather than overestimate drug risk. Our use of other diabetes medications as comparators, however, should reduce or eliminate the majority of these potential biases, although we cannot fully exclude biases introduced by physicians or patients leading to selection of specific drugs. Furthermore, there may be increased cardiovascular risk with rosiglitazone for patients using insulin, which may also be a surrogate for duration of diabetes. In addition to adjusting for insulin in our models, we performed stratification, yielding very similar results. Notably, no patients receiving pioglitazone without additional outpatient insulin were identified to have a myocardial infarction. Future analyses should consider drug combinations because concomitant use of insulin and thiazolidinediones may be particularly unfavorable (10,21). Furthermore, our low specificity for detection of myocardial infarction events of 74% is of particular concern, indicating a need for future analyses to incorporate laboratory data to verify the occurrence of myocardial infarction more accurately. Composite end points of major adverse cardiovascular events are standard measures for comparing treatments in large cardiovascular outcome studies. Our analysis included only myocardial infarction, whereas other cardiovascular events, such as sudden death and stroke, were not considered in this analysis. Finally, if there is increased health risk shortly after initiation of therapy that is abrogated with longer duration of administration, then cumulative assessments including additional new patients and patient-year exposure would tend to produce a bias toward early risk.

The control of residual confounders in observational data is an important issue. Approaches addressing this issue in medical record data include comparing risk in groups for the measured outcome before and after an exposure (23) to test whether a group was at prior higher risk, picking comparable exposures (medications in the same class) where heuristically there is no reasonable argument for differences in groups, and using global, accepted measurements of acuity (such as the Charlson score) to detect differences in underlying health of groups. We selected the latter method, since there was some suggestion of risk differences for the two marketed thiazolidinedione products available for study.

Although the increased risk ratio for rosiglitazone compared with other diabetes medications has been demonstrated in more robust clinical datasets with adequate longitudinal records of patients (21), the current study provides two novel and important insights. First, with the need to monitor numerous products and numerous potential events, it is increasingly difficult to develop randomized clinical trials to adequately address all potential study bias and confounding factors. From a surveillance perspective, a real-time strategy detecting risk that may require further investigation is potentially more cost-effective than numerous longterm investigations into one drug-one event relationships. Moreover, designing studies to identify relatively infrequent, but medically important, adverse events that would probably be missed by phase III clinical trials and current postmarketing voluntary reporting mechanisms would probably be expensive and curtail or delay development of new treatments. Surveillance analysis should be guided by a priori evidence (such as nonstatistically significant adverse events) from phase III clinical trials to limit the potential for false-positive results. It is important to note that surveillance methods work best when agents have adequate population uptake. For instance, the time to identify signals comparing rosiglitazone and pioglitazone was delayed because of low sample sizes for both drugs. Methodologies that improve detection performance, especially when drugs and events are rare, or permit all possible drug-event interactions are described elsewhere (24).

Second, our study shows how relatively simple clinical surveillance methods can be implemented in real time. With the availability of electronic datasets such as the one used herein, it is possible to perform analyses of drug-event combinations prospectively on a quarterly, monthly, or even weekly basis. In this study, we demonstrated that if these methods had been in use when thiazolidinediones were first introduced to the market, a potential hazard would have been apparent ~18 months after the launch, in 2001, well before concerns were raised publicly in 2007 (3). This time frame is also faster than would be realized in phase IV postmarketing trials and may cause less delay than requiring cardiovascular outcome trials before FDA approval for diabetes medications that do not have adverse safety signals in aggregate phase II–III study analysis. Although these methods would not provide the same degree of information as a prospective randomized control trial, they might indicate caution to care providers faced with options to prescribe multiple newer medications, fulfilling clear needs for complementary approaches (25).

Our study provides a framework for implementation of future postmarketing surveillance activities with semiautomated extraction of large clinical datasets. Despite inherent limitations, these data can provide robust real-time signals of adverse drug events in the postmarketing setting. How such systems will interact with activities at the FDA requires thoughtful consideration.

Acknowledgments— This work was supported in part by the National Institutes of Health National Center for Biomedical Computing (Grant 5U54-LM-008748).

No potential conflicts of interest relevant to this article were reported.

### References

- Misbin RI. Lessons from the Avandia controversy: a new paradigm for the development of drugs to treat type 2 diabetes. Diabetes Care 2007;30:3141–3144
- Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay DK, Platt R, Black S, Shinefield H, Chen RT. Active surveillance of vaccine safety: a system to detect early signs of adverse events. Epidemiology 2005;16:336–341
- Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med 2007;356:2457–2471
- Singh S, Loke YK, Furberg CD. Long-term risk of cardiovascular events with rosiglitazone: a meta-analysis. JAMA 2007;298: 1189–1195
- Home PD, Pocock SJ, Beck-Nielsen H, Curtis PS, Gomis R, Hanefeld M, Jones NP, Komajda M, McMurray JJ, RECORD Study Team. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised,

open-label trial. Lancet 2009;373:2125-2135

- Walker AM, McAfee AT, Koro C. Studies of diabetes, thiazolidinediones, and coronary heart disease. Pharmacoepidemiol Drug Saf 2007;16:1313–1314; author reply 1314–1316
- Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. Am Heart J 2004;148:99–104
- Haffner SM, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med 1998;339:229–234
- Psaty BM, Furberg CD. Rosiglitazone and cardiovascular risk. N Engl J Med 2007; 356:2522–2524
- Food and Drug Administration. Advisory Committee meeting on July 30th, 2007, to discuss cardiovascular ischemic events with Avandia [article online], 2007. Available from www.fda.gov/ohrms/dockets/ac/07/briefing/2007-4308b1-02-FDA-backgrounder.pdf. Accessed 4 November 2009
- Diamond GA, Bax L, Kaul S. Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. Ann Intern Med 2007;147:578– 581
- Kahn SE, Haffner SM, Heise MA, Herman WH, Holman RR, Jones NP, Kravitz BG, Lachin JM, O'Neill MC, Zinman B, Viberti G, ADOPT Study Group. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. N Engl J Med 2006; 355:2427–2443
- Home PD, Pocock SJ, Beck-Nielsen H, Gomis R, Hanefeld M, Jones NP, Komajda M, McMurray JJ, RECORD Study Group. Rosiglitazone evaluated for cardiovascular outcomes—an interim analysis. N Engl J Med 2007;357:28–38
- 14. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC Jr, Bigger JT, Buse JB, Cushman WC, Genuth S, Ismail-Beigi F, Grimm RH Jr, Probstfield JL, Simons-Morton DG, Friedewald WT. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;358: 2545–2559
- 15. BARI 2D Study Group, Frye RL, August P, Brooks MM, Hardison RM, Kelsey SF, MacGregor JM, Orchard TJ, Chaitman BR, Genuth SM, Goldberg SH, Hlatky MA, Jones TL, Molitch ME, Nesto RW, Sako EY, Sobel BE. A randomized trial of therapies for type 2 diabetes and coronary

artery disease. N Engl J Med 2009;360: 2503–2515

- 16. Dormandy JA, Charbonnel B, Eckland DJ, Erdmann E, Massi-Benedetti M, Moules IK, Skene AM, Tan MH, Lefèbvre PJ, Murray GD, Standl E, Wilcox RG, Wilhelmsen L, Betteridge J, Birkeland K, Golay A, Heine RJ, Korányi L, Laakso M, Mokán M, Norkus A, Pirags V, Podar T, Scheen A, Scherbaum W, Schernthaner G, Schmitz O, Skrha J, Smith U, Taton J, PROactive Investigators. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial. Lancet 2005;366:1279-1289
- 17. Lincoff AM, Wolski K, Nicholls SJ, Nissen SE. Pioglitazone and risk of cardiovascular events in patients with type 2 diabetes mellitus: a meta-analysis of randomized trials. JAMA 2007;298:1180–1188
- Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 2008;359:1577–1589
- Gerrits CM, Bhattacharya M, Manthena S, Baran R, Perez A, Kupfer S. A comparison of pioglitazone and rosiglitazone for hospitalization for acute myocardial infarction in type 2 diabetes. Pharmacoepidemiol Drug Saf 2007;16:1065–1071
- 20. Winkelmayer WC, Setoguchi S, Levin R, Solomon DH. Comparison of cardiovascular outcomes in elderly patients with diabetes who initiated rosiglitazone vs pioglitazone therapy. Arch Intern Med 2008;168:2368–2375
- Walker AM, Koro CE, Landon J. Coronary heart disease outcomes in patients receiving antidiabetic agents in the Phar-Metrics database 2000–2007. Pharmacoepidemiol Drug Saf 2008;17:760–768
- Lipscombe LL, Gomes T, Lévesque LE, Hux JE, Juurlink DN, Alter DA. Thiazolidinediones and cardiovascular outcomes in older patients with diabetes. JAMA 2007;298:2634–2643
- 23. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. BMJ 2009;338:b81
- 24. Chan KA, Hauben M. Signal detection in pharmacovigilance: empirical evaluation of data mining tools. Pharmacoepidemiol Drug Saf 2005;14:597–599
- 25. Avorn J. In defense of pharmacoepidemiology—embracing the yin and yang of drug research. N Engl J Med 2007;357: 2219–2221

# The Hub Population Health System: distributed ad hoc queries and alerts

Michael D Buck,<sup>1,2</sup> Sheila Anane,<sup>1</sup> John Taverna,<sup>1</sup> Sam Amirfar,<sup>1</sup> Remle Stubbs-Dame,<sup>1</sup> Jesse Singer<sup>1</sup>

# ABSTRACT

<sup>1</sup>Bureau of Primary Care Information Project, New York City Department of Health and Mental Hygiene, New York, New York, USA <sup>2</sup>Department of Biomedical Informatics, Columbia University, New York, New York, USA

#### Correspondence to

Dr Michael D Buck, New York City Department of Health and Mental Hygiene, 42-09 28th Street — CN#52, Queens, NY 11101-4132, USA; mbuck@health.nyc.gov

Received 20 April 2011 Accepted 16 October 2011 The Hub Population Health System enables the creation and distribution of queries for aggregate count information, clinical decision support alerts at the pointof-care for patients who meet specified conditions, and secure messages sent directly to provider electronic health record (EHR) inboxes. Using a metronidazole medication recall, the New York City Department of Health was able to determine the number of affected patients and message providers, and distribute an alert to participating practices. As of September 2011, the system is live in 400 practices and within a year will have over 532 practices with 2500 providers, representing over 2.5 million New Yorkers. The Hub can help public health experts to evaluate population health and quality improvement activities throughout the ambulatory care network. Multiple EHR vendors are building these features in partnership with the department's regional extension center in anticipation of new meaningful use requirements.

### BACKGROUND

Traditionally, public health departments have gathered information concerning infectious disease outbreaks through mandated provider reporting and chronic disease burden through in-depth interviews of afflicted patients. These methods are costly and resource intensive (eg, the Health and Nutrition Examination Surveys) and require years of planning and execution to carry out.<sup>1 2</sup> Additionally, reportable disease registries, birth/death statistics, syndromic surveillance, immunization reporting, etc, tend to be stored in specific program areas, making them difficult to use across a health department.<sup>3–8</sup> These approaches cannot respond quickly enough to evolving public health priorities and new discoveries in healthcare. They do not provide a 'real-time' assessment of the comprehensive needs of a community and are therefore difficult to use for program planning and resource allocation. More innovative approaches are needed that focus on integrating public health priorities directly into the clinical provider's workflow using the electronic health record (EHR).<sup>9 10</sup>

In order to develop a system that will allow the flexibility to investigate and monitor unanticipated acute events, respond to changing public health needs around chronic disease burdens, and identify causative agents in a timely manner, the New York City Department of Health and Mental Hygiene (DOHMH) has worked to develop a public healthoriented EHR. Combining the EHR with ad hoc query and alert features would allow public health officials to quickly determine disease burden or investigate outbreaks by geographic zip code, comorbidities, race/ethnicity, etc, and give them a clinical action arm to influence chronic disease care services throughout the community. Furthermore, combining clinical information with additional city health data sources could form the basis for a population heath record for monitoring community health as envisioned by the AMIA Board of Directors in 1997 and reiterated by Friedman *et al* in 2010.<sup>11</sup> <sup>12</sup>

### METHODS Setting

### The Primary Care Information Project (PCIP) is a bureau within the DOHMH formed in 2005 with the mission to develop and implement a public health-enabled EHR in ambulatory primary care practices serving the medically underserved.<sup>13</sup> <sup>14</sup> This EHR includes an integrated decision support system based upon standardized clinical quality measure standards, real-time alerts, order sets, quality measure dashboards, ad hoc registry queries, and health information exchange capabilities.

Through a competitive procurement process, eClinicalWorks was chosen as the initial EHR vendor partner to develop these capabilities and deploy them to all participating practices. As of July 2011, PCIP has implemented the EHR in 532 practices whose 2506 providers serve an estimated 2.5 million patients in 4.8 million encounters per year. These practices range in size from solo practitioners to large, multi-site, 100+ provider community health centers. As PCIP has begun expanding its operations to include regional extension center activities, additional EHR vendor partners and their associated practices are also being added to the program.

### System overview

The Hub Population Health System (Hub) was built as a joint collaboration between PCIP and eClinicalWorks beginning in November 2009. Unlike large integrated healthcare delivery systems, PCIP practices are part of a 'virtual network' of distributed independent ambulatory practices (see figure 1). Each individual EHR clinical data repository connects on a nightly basis to a central server (the Hub), hosted by the vendor, to receive and transmit information using a secured HTTPS connection. All information is summarized at the aggregate count level before transmission to the Hub. This helps to protect patient privacy by limiting the information shared between

### **Brief communication**

**Figure 1** Hub system architecture for a virtual network of practices.



institutions, a strategy which has been described in depth by other researchers.<sup>15-17</sup> Data from the Hub are downloaded nightly to a secured data warehouse hosted by PCIP. All practices sign data sharing agreements which permit the sharing and use of this aggregate data with PCIP. No aggregate data with practice identifiers are shared with third parties unless specifically authorized to do so by practices.

The Hub Population Health System provides four primary services to authorized users. First, it permits the distribution of SQL query reports for aggregate count information and EHR point-of-care decision support alerts. Second, it enables the distribution of the reports/alerts to any practice in the network according to defined reporting policies. Third, it provides an interface for viewing and downloading aggregate results reported from the queries run on each of the practices. Fourth, it has the ability to securely message providers directly in their EHR inbox.

As of September 2011, there were 400 Hub-enabled practices, covering 1.6 million patients. A total of 736 unique queries have returned 220 406 results from these practices. The maximum number of queries run by one practice in a single day was 177. A number of error conditions have been reported, including practice installation/configuration issues, MySQL/SQL Server SQL query incompatibilities, and missing transmissions, particularly on weekends/holidays.

### Query/alert building

Figure 2 shows the primary navigation panel for the Hub. It illustrates the creation of a new policy with metronidazole reports and a decision support alert. Each report requires a name, description, SQL query code, and decision support flag. The name uniquely identifies each report and serves as the primary alert text (eg, DOH Alert for Metronidazole. Click "?" for details) when the decision support flag is enabled. The description provides additional details to the clinician about the query/alert and an email address for follow-up questions.

The SQL query code is generated by public health domain experts leveraging the EHR's Registry reporting tool which has a GUI interface for entering patient characteristics including demographics, medications, diagnoses, etc. The output of this tool is automatically extracted and formatted to be compatible with the Hub using a PCIP-built JAVA plugin called the 'Phactory'. The combination of these tools creates a process for Hub users to create a wide range of queries with no SQL programming expertise required. More advanced queries are manually written by PCIP analysts; however, future enhancements to the Phactory will automate much of this work.

### **Policy distribution**

Each report is attached to a policy which contains all the reporting metadata. A policy identifies which practices will run the report, the policy start and end dates, the frequency at which the results will be uploaded to the central hub, the report start and end dates, and report frequency. In figure 2, the policy's reports will upload results every day from January 12 to January 15.

On a nightly basis, each practice's EHR system runs a scheduled job that connects over HTTPS to the central Hub server hosted by eClinicalWorks (see figure 1). New policies are downloaded and executed against the local practice's EHR database, and the results are transmitted back to the Hub server. If a policy has expired, it is marked as inactive in the local practice database. Errors in query execution are automatically sent via email to a PCIP administrative account for corrective follow-up action.

All currently active policies for a given practice are displayed in the Registry Reports section of their EHR. A provider can read the report description and email the Hub administrator with any follow-up questions. If an alert has been assigned, it will appear on the right panel of the screen when the provider is documenting in the progress note of a patient meeting the criteria. The alert text can include follow-up links to relevant provider/patient educational material, disease reporting websites, provider surveys, etc, based upon the suggested intervention. The alert will disappear once the appropriate follow-up action is taken (eg, ordering a laboratory test) or the alert's policy expires. Currently the alerts are text-based, but future enhancements will allow direct action through computerized physician order entry.

PHA Administation 🗲	Define Ne	w Policy						
Policy Management		📥 Add P	ractice 😂 Delete	📥 Add Report 🛛 😂 Delete				
	Practice Cod	Practice Name	JoinDate	Report Name	CDSS			
Compose Message	5513	eCWUppradeTesting80	10/13/2010	DOH Metronidazole 1 20101213 to 20110113	No			
PHA Reports				DOH Alert for Metronidazole. Click "?" for details.	Yes			
Query Builder								
Query Management		Please enter v	alue in all the required(*) f	ields.				
Message Log	Policy Star	Policy Start Date: * 01/12/2011						
	Policy En	d Date: * 01/15/2011	<b></b>					
	Upload Freq	uency: * 1	Days O Week	s 🔾 Months 🔾 Years				
	Report Star	t Date: * 12/13/2010						
	Reporting I	Period: * 1	🔵 Days 🔵 Week	is 💿 Months 🔾 Years				
	Report En	d Date: 01/13/2011						
	Report Freq	uency: * 1	Days O Week	s 🔾 Months 🔾 Years				

Figure 2 Distribution of a new policy with metronidazole reports and alert (eClinicalWorks).

### **Result reporting**

The Reports section contains all of the results from policies which have been assigned. There are options to filter the data by practice, report name, and reporting date. This information is exported as an Excel file for upload to the PCIP data warehouse for further analytic work. This file contains a unique practice identifier, practice name, report identifier, report name, report run date, report period start and end dates, and the report count. Future enhancements to the system will include automated exporting of the data via web services.

### **Provider messaging**

Finally, the Hub system permits secure messaging of targeted practices using HTML-formatted messages. Messages are delivered to the primary practice provider's inbox used for secure internal practice communication. Each message can be assigned a priority which marks the message as Routine, Urgent, and Emergent. A Message Log page is used to track all messages transmitted to the practices.

### **PILOT EVALUATIONS**

### Alert campaigns for public health emergencies

Since we can tailor the distribution and timing of the alerts to specific practices for specific times, we can conduct targeted 'alert campaigns' that will have optimal clinical impact while minimizing the potential for alert fatigue, which has been studied previously.<sup>9</sup> <sup>18</sup> We tested such a campaign using a drug recall issued by the CDC for metronidazole, an antibiotic commonly used to treat specific infections in the outpatient setting.

On January 6, 2011, the FDA recalled underweight metronidazole tablets.<sup>19</sup> On January 12, the DOHMH Health Alert Network distributed a clinical notification describing the recall.<sup>20</sup> Using this recall notification, clinical and public health experts distributed queries to two pilot practices that same evening. The reports returned a count of 62 patients in six stratified time periods who were prescribed metronidazole in the last year. A secure follow-up message was sent on January 14 to the providers' inboxes embedded in the EHR that included the specifics of the recall, as well as step-by-step instructions on how to use the EHR's Registry function to identify the affected patients for purposes of patient notification. The message also included a hyperlink to the FDA's MedWatch website with detailed information on the recall.

After reviewing the data, a clinical decision support alert entitled 'DOH Alert for Metronidazole' was activated from January 19, 2011 to February 19, 2011. For any patient prescribed metronidazole in these two practices in the last 60 days, the alert appeared in the right pane of the progress note documentation screen. The pop-up information window contained the text of the recall, pertinent hyperlinks, and a reminder to review the longer inbox message.

After the message and alert were distributed, we spoke with three providers to understand the caregiver's perspective. Two providers had read the message and used the hyperlinks to read further information about the recall. Also two of the providers saw the progress note alert and agreed that it was useful and informative. The third provider, who had not seen the alert, agreed that receiving these types of interventions would be useful once a week and during an emergency. One provider agreed strongly that the alert changed the way he/she practiced medicine in reacting to this recall. Although this application of queries and alerts was extremely exploratory, it suggests the potential of this interactive technology to improve the way health departments gauge the severity of public health emergencies and connect with providers to deal with them.

### **Population health analysis**

The query feature of the Hub can also be used to examine the distribution of disease on a population level and to inform program planning in real time. As an example, we probed the utility of aggregate count data for public health by looking at neighborhood-level diabetes prevalence among PCIP patients. Our denominator was patients seen in 2010 with a standard

### **Brief communication**

NYC zip code, organized into United Hospital Fund (UHF) neighborhoods. Our numerator was the subset of those patients with an ICD-9 code for diabetes on the problem list in 2010.

From August 22 to September 12, 2011, we queried 386 small practices for the numerator and denominator in each UHF neighborhood. These queries returned 28993 distinct results representing 4 49 775 patients at 381 responder practices, 43 423 of whom had a diagnosis of diabetes. There were an additional 158 error values. To deal with missing values, we allowed only practices with a complete pair of numerator and denominator data for a given neighborhood to contribute to the prevalence estimate for that neighborhood. We transformed the data into numerator/denominator variables, with one observation per practice, aggregating them into neighborhood-level counts and mapping that prevalence (figure 3). While this form of analysis does not achieve the results that more sophisticated multi-level modeling may yield, aggregate geographic information like this is both applicable and essential to health departments trying to target resources where they are most needed.

### DISCUSSION

By the first quarter of 2012, the Hub system will cover nearly 2.5 million New Yorkers, or almost 30% of the 8.4 million residents in NYC. PCIP is working to extend the Hub to additional EHR

2010 diabetes prevalence among NYC patients

by UHF Neighborhood



Practice 7	15	54	82	140	226	919	575	2
Practice 6	31	20	23	34	20	23	7	33
Practice 5	0	0	0	0	1	1	0	1733
Practice 4	0	2	3	1	1	1	2	18
Practice 3	6	5	12	19	4	9	4	11
Practice 2	46	43	68	118	38	57	32	95

N=449 775 patients (43 423 diabetic patients) with standard zip code recorded at 381 practices

**Figure 3** Illustration of geographic population health analysis with the Hub.

vendor partners, which will enable greater coverage of the NYC ambulatory patient population. This is a significant patient population in which to monitor healthcare outcomes and intervene in near real-time on critical priorities using public health alerts. Through the Hub system, the Health Department can investigate population health issues without clinical or vendor resources. This type of innovation has already had national implications—similar alerts are being considered as a required EHR feature for Meaningful Use Stage 3.<sup>21</sup>

The datasets we derive from the Hub can currently be used with classic regression techniques to examine quality of care at the practice level. More research is needed to adapt existing multi-level modeling of fixed and random effects to examine patient-level phenomena. Given the geographic component of these queries, these EHR datasets can be linked to other GIS data like air quality and census socioeconomic information to give a more complete picture of health issues and disparities throughout NYC. Eventually these de-identified datasets may even form the basis for an aggregate population health record (popHR) for monitoring health citywide.<sup>12</sup> As Hub-like system features, such as ONC's Query Health initiative, penetrate other communities in the nationwide EHR marketplace, they could be used in conjunction with existing public health systems to form a nationwide strategy for population health monitoring and research.<sup>10</sup> <sup>16</sup> <sup>22–26</sup>

One important consideration of our approach is that it requires each EHR vendor to provide a separate proprietary interface for creating and sharing public health queries among practices with the same system. The alternative approach would require each EHR vendor to agree to a standardized mapping of their data to a well-defined, uniform subset.<sup>24 26</sup> Each approach has its own limitations that should be weighed against the needs and capabilities of local public health and clinical organizations. Other potential limitations in using the Hub system to monitor population health are that the doctorgoing population may not be representative of the general population and it has no way to eliminate duplicate patient counts across multiple distinct practices. More in-depth evaluations of the system to compare the Hub data to patient surveys and chart reviews as a formal validation of its quality and reliability are needed.

Acknowledgments We would like to thank the developers and management of eClinicalWorks for their support of this project.

 $\ensuremath{\textit{Funding}}$  This project is funded by the New York City Department of Health and Mental Hygiene.

#### Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

### REFERENCES

- Thorpe LE, Gwynn RC, Mandel-Ricci J, et al. Study design and participation rates of the New York City Health and Nutrition Examination Survey, 2004. Prev Chronic Dis 2006;3:A94.
- Gwynn RC, Berger M, Garg RK, et al. Measures of adiposity and cardiovascular disease risk factors, New York City Health and Nutrition Examination Survey, 2004. Prev Chronic Dis 2011;8:A56.
- Hopkins RS. Design and operation of state and local infectious disease surveillance systems. J Public Health Manag Pract 2005;11:184–90.
- Dato V, Wagner MM, Fapohunda A. How outbreaks of infectious disease are detected: a review of surveillance systems and outbreaks. *Public Health Rep* 2004;119:464-71.
- Loonsk JW. BioSense—a national initiative for early detection and quantification of public health emergencies. *MMWR Morb Mortal Wkly Rep* 2004;53(Suppl): 53-5.
- 6. Epiquery. http://www.nyc.gov/health/epiquery (accessed 31 Mar 2011).
- Heffernan R, Mostashari F, Das D, et al. New York City syndromic surveillance systems. MMWR Morb Mortal Wkly Rep 2004;53(Suppl):23–7.
- Chamany S, Silver LD, Bassett MD, *et al.* Tracking diabetes: New York City's A1C Registry. *Milbank Q* 2009;87:547–70.
- Lurio J, Morrison FP, Pichardo M, et al. Using electronic health record alerts to provide public health situational awareness to clinicians. J Am Med Inform Assoc 2010;17:217–19.
- Klompas M, Lazarus R, Daniel J, et al. Automated public health reporting-possible with a coalition of the willing. Advances in Disease Surveillance 2007;395:1–2.
- 11. **Detmer DE.** Activating a full architectural model: improving health through robust population health records. *J Am Med Inform Assoc* 2010;**17**:367–9.
- Friedman DJ, Parrish RG 2nd. The population health record: concepts, definition, design, and implementation. J Am Med Inform Assoc 2010;17:359-66.
- Primary Care Information Project. http://www.nyc.gov/html/doh/html/pcip/pcip.shtml (accessed 30 Mar 2011).
- New York City Regional Electronic Adoption Center for Health. http://www.nycreach. org (accessed 31 Mar 2011).
- Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. *Health Aff (Millwood)* 2009;28:454–66.
- Maro JC, Platt R, Holmes JH, *et al.* Design of a national distributed health data network. *Ann Intern Med* 2009;151:341–4.
- Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. BMC Public Health 2006;6:235.

- Weingart SN, Toth M, Sands DZ, et al. Physicians' decisions to override computerized drug alerts in primary care. Arch Intern Med 2003;163:2625–31.
- U.S. Food and Drug Administration. http://www.fda.gov/Safety/MedWatch/ SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm239312.htm (accessed 5 Apr 2011).
- New York City's DOHMH Health Alert Network. https://a816-healthpsi.nyc.gov/han/ nychan/ (accessed 5 Apr 2011).
- The Office of the National Coordinator for Health Information Technology. http://healthit. hhs.gov/media/faca/MU RFC%20 2011-01-12 final.pdf (accessed 31 Mar 2011).
- Friedman DJ. Envisioning futures for state Web-based data query systems. J Public Health Manag Pract 2006;12:196–200.
- Friedman DJ. Assessing the potential of national strategies for electronic health records for population health monitoring and research. National Center for Health Statistics. *Vital Health Stat 2* 2006;143:1–83.
- Miller C. CDC to test actionable alerts with GE Healthcare [press release]. Orlando: GE Healthcare, 2011.
- Rajeev D, Staes CJ, Evans RS, et al. Development of an electronic public health case report using HL7 v2.5 to meet public health needs. J Am Med Inform Assoc 2010;17:34–41.
- Standards & Interoperability (S&I) Framework Query Health. http://wiki. siframework.org/Query+Health (accessed 3 Oct 2011).

### Electronic Laboratory Data Quality and the Value of a Health Information Exchange to Support Public Health Reporting Processes Brian E. Dixon, MPA, PhD<sup>1,2</sup>, Julie J. McGowan, PhD<sup>1,2,3</sup>, Shaun J. Grannis, MD, MS<sup>1,2,3</sup>

#### <sup>1</sup>Indiana University, School of Informatics, Indianapolis, IN; <sup>2</sup>Regenstrief Institute, Indianapolis, IN; <sup>3</sup>Indiana University, School of Medicine, Indianapolis, IN

#### Abstract

There is increasing interest in leveraging electronic health data across disparate sources for a variety of uses. A fallacy often held by data consumers is that clinical data quality is homogeneous across sources. We examined one attribute of data quality, completeness, in the context of electronic laboratory reporting of notifiable disease information. We evaluated 7.5 million laboratory reports from clinical information systems for their completeness with respect to data needed for public health reporting processes. We also examined the impact of health information exchange (HIE) enhancement methods that attempt to improve completeness. The laboratory data were heterogeneous in their completeness. Fields identifying the patient and test results were usually complete. Fields containing patient demographics, patient contact information, and provider contact information were suboptimal. Data processed by the HIE were often more complete, suggesting that HIEs can support improvements to existing public health reporting processes.

#### Introduction

Interest in and development of methods for leveraging electronic health record (EHR) data across disparate sources for a variety of use cases (a.k.a. secondary use) is widespread (1). Existing literature discusses the challenges associated with the re-use of EHR data for purposes beyond clinical care, including access to data, privacy protections, identity matching, and the interoperability of data from disparate sources (1, 2). However, often missing from discussions of secondary use is a common, core issue that arguably is more problematic than any other issue relevant to re-use of EHR data: poor data quality.

Poor data quality is common and affects all industries and organizations that employ information systems (3). Typical data quality issues encountered include: inaccurate data, inconsistencies across data sources, and incomplete (or unavailable) data necessary for operations or decisions (4). A large bank found that data in its credit-risk management database were only 60% complete, which necessitated additional scrutiny by anyone using its data (5). In health care, the completeness of data in EHR systems has been found to vary from 30.7 to 100% (6).

Good evidence from the information management literature on the impacts of these issues is sparse, but good estimates of impacts include: increased costs in the range of 8-12% of organizational revenue, and up to 40-60% of a service organization's expenses consumed as a result of poor data; poorer decisions that take longer to make; lower data consumer satisfaction with information systems; and increased difficulty in reengineering work and information flows to improve service delivery (4). Impacts on health care include poorer decisions when humans or machines use poor quality data inputs from EHR systems (7, 8).

Data quality issues have been well examined and documented in the epidemiology literature. For example, spontaneous reporting rates for infectious diseases range from 9% to 99% and have remained relatively unchanged from 1970 - 2000 (9). While some conditions, such as sexually transmitted infections, are reported approximately 80% of the time, many conditions are reported less than half of the time. Timeliness, another attribute of data quality (3), has also been found to be a challenge in public health reporting (10). Delays in the receipt of notifiable disease data (timeliness) and the lack of a complete set of reports (completeness) impact public health agency surveillance processes, including but not limited to the ability of agencies to respond to emerging disease threats.

Electronic laboratory reporting (ELR) was demonstrated just over a decade ago to be an effective method to improve the timeliness of reporting as well as the number of reports submitted to public health agencies (10). Since ELR was shown to be effective, the U.S. government, states, and a number of private foundations have invested millions of dollars into the development, implementation, and adoption of health information, ELR, and surveillance systems (11-15). Despite reported improvements in the timeliness and volume of submitted reports, some studies indicate anecdotally that ELR may not improve the completeness of the data in the submitted reports (16). Given a paucity of evidence in the literature that ELR does or does not impact the completeness of notifiable disease data, this study examined the completeness of data from clinical information systems. In addition to characterizing the completeness of ELR data, the study further compared raw data directly sent from clinical information systems with data enhanced by a health information exchange (HIE) prior to transmission to a public health agency. If an HIE can improve the completeness of ELR data submitted to public health, it would signify that HIE data enhancement methods are a valuable, effective method for improving notifiable disease data quality. Improving data quality will likely translate into improvements in disease surveillance processes, impacting both clinicians and public health professionals.

#### Background

Although surveillance methods and practices date back to 1854 when John Snow used reported mortality data and location information to convince authorities to remove a water pump that was the source of a cholera outbreak, modern surveillance activities are aided by computer systems and informatics methods (17). This modernization of surveillance began in earnest during the previous decade following a report commissioned by the U.S. Department of Health and Human Services (HHS) and U.S. Centers for Disease Control and Prevention (CDC) (18) as well as early evidence published on the use of electronic methods to enhance traditionally manual surveillance processes (10). The HHS report and early pioneers illuminated a number of challenges, including but not limited to: 1) a lack of data standards for the exchange of surveillance data between providers and public health and between public health entities; 2) variability in the use of available messaging formats for the exchange of surveillance data between providers and public health and between public health entities; 3) limited decision and analytic support from early computer applications; and 4) a general lack of computer systems in public health laboratories (18, 19).

Policymakers responded to these challenges by funding numerous initiatives which included at least one principal aim to modernize surveillance practices by implementing advanced IT systems and networks that would link health care providers and public health agencies together to better detect and cooperatively address disease outbreaks (15, 20). Funding from the CDC Office of Surveillance, Epidemiology and Laboratory Services is just one example (21).

Many initiatives focused on electronic laboratory reporting (ELR), which involves the electronic submission of laboratory data, following the confirmation of an infectious disease, to a public health agency. Others focused on syndromic surveillance which detects initial manifestations of disease before clinical or laboratory diagnoses are established. All initiatives sought to improve the timeliness, accuracy, and completeness of data needed by public health agencies to perform surveillance activities.

The evidence in recent literature demonstrates that ELR can be effective at improving the timeliness of infectious disease reports (22), and ELR can further increase the number of cases reported to public health agencies (23). However, some researchers anecdotally suggest that ELR may not improve the completeness of the data reported to public health agencies (16). Therefore our study focused on evaluating the completeness of the data output from current clinical information systems. We examined ELR data received directly from laboratory and hospital information systems as well as enhanced data transmitted from an HIE to public health. For the purposes of this study, enhanced data is defined as data that has been validated, including but not limited to correcting units of measure, as well as augmented, including but not limited to mapping local lab test identifiers to standardized vocabulary concepts. Researchers at Regenstrief have previously described the practice of data enhancement and its routine use when exchanging clinical data as a method to improve data quality and interoperability (24, 25).

#### Methods

The scope of our research included the following aims: 1) the development of a method for evaluating the completeness of laboratory data in the context of public health reporting; 2) measuring the completeness of laboratory data received from clinical information systems and an HIE using the method; and 3) comparing the completeness of the "raw" data from clinical information systems (e.g., unaltered, unedited ELR messages) with the completeness of "enhanced" data from the HIE (e.g., ELR messages having syntax corrected and concepts mapped to standard vocabularies). The study was performed in the context of the lead author's (BED) dissertation and approved by both the Indiana University-Purdue University Indianapolis Institutional Review Board as well as the research council of the Indiana Network for Patient Care (INPC).

The central theme of the study was completeness. Completeness in the context of public health surveillance refers to both the proportion of diagnosed cases reported to public health and the proportion of fields in a case report completed by the submitting hospital or lab (26). As previously described, ELR messages' ability to increase the proportion of diagnosed cases reported to public health has been well established (10, 16, 22). Therefore this study

concentrated on measuring the completeness of the data within ELR messages transmitted by a data source (e.g., hospital, laboratory, HIE). The completeness of ELR messages in this context is unknown and only asserted anecdotally by previous research in public health informatics.

A first step in analyzing the completeness of ELR message data involved creating a "minimum data set" for ELR messages that would meet the information needs of public health agencies that receive the data. Public health professionals collate ELR data with data received from other sources to complete a case report. Therefore the aim of ELR should be to provide a set of data that can populate as much of the case report as possible to streamline public health and clinical workflows. For example, when data is missing, public health professionals often call provider organizations to acquire the missing data, which disrupts public health and clinical workflows.

The lead author (BED) created the "minimum data set" by identifying the data elements required under Indiana law for laboratories to report to public health agencies. This initial set was then augmented using data elements reported in the public health literature to be useful to public health agencies for surveillance activities. This refined set was then provided to public health professionals and researchers for review and comment. Feedback was used to create a final list of data elements that would profile a given ELR message set's completeness.

Next a "completeness profile" was calculated for two samples of real-world ELR messages using the minimum data set. Each profile was constructed by dividing the number of values present in a given field by the total possible values that could have been populated in that field. The profiles were then compared to one another.

The study data originated from production information systems utilized by a variety of clinical settings and an HIE. The first sample contained "raw" (unaltered, unedited) Health Level 7 (HL7) messages (Version 2.x) received from 168 distinct hospital and laboratory information system interfaces during a one month period (November 14, 2010 to December 15, 2010) by the INPC, an operational HIE (27) that includes integrated delivery networks, hospitals, independent laboratories, physician practices, radiology centers, and the Indiana State Department of Health (ISDH). These messages were extracted from the INPC's inbound message queue and parsed into a relational database composed of tables that represent logical HL7 segments (e.g., PID, OBR, OBX). Each field within the tables corresponded to an individual field within a HL7 segment (e.g., PID-1, OBR-16, OBX-3). To parse the messages, we employed a clone of the INPC's production methods for deconstructing HL7 messages. These methods are used to receive and process real-world HL7 messages from a variety of clinical information systems, and they have been refined and validated over the INPC's 16 years of operation.

The second sample contained "enhanced" HL7 messages (Version 2.x) representing 49 distinct hospital and laboratory sources processed by the INPC during the same timeframe. These messages were extracted from the outbound message queue which contains reportable messages bound for the state public health agency. The INPC utilizes the Regenstrief Notifiable Condition Detector (NCD) to critically examine HL7 messages from INPC interfaces that potentially contain notifiable disease results. Messages determined to contain reportable results are sent from the outbound queue to the ISDH on behalf of the INPC and its member institutions (e.g., hospitals, labs).

The NCD further enhances the HL7 messages through validation and augmentation methods. For example, local laboratory codes contained within the OBX-3 field are mapped to equivalent Logical Observation Identifiers Names and Codes (LOINC) codes. The LOINC codes are appended to the original messages prior to transmission to the ISDH. The NCD further examines incoming messages for provider information (e.g., National Provider Identifier, address of the hospital or practice, phone number for the department or clinic) and attempts to add any missing provider information found in a table of providers stored in the INPC knowledge repository. Furthermore, labs may improperly place units of measure in a comment field, so the NCD examines comments for key data like units of measure and copies relevant data it finds to the appropriate HL7 field. The messages in this sample were parsed using the same methods as the raw sample into the same relational database for analysis.

Structured query language (SQL) statements were executed to calculate the completeness of each HL7 field within both samples. Each field's "percent complete" was calculated by dividing the count of non-null values by the total number of possible values for that field. The calculated values were input into a completeness profile for each sample, and the difference between the completeness scores across samples was calculated.

#### Results

#### Aim 1: Key Fields that Support Notifiable Disease Surveillance Processes

The result of the first aim in this study was the development of a novel method for measuring completeness and comparing the completeness of two or more data sets. We first defined a "minimum data set" that contains the data

elements specifically required in state law that are to be reported to public health agencies for notifiable conditions. In Indiana, these data elements are defined in the Indiana Administrative Code (IAC) under 410 IAC 1-2.3-48.

In addition to what is minimally required by law, the minimum data set was constructed to also include those additional elements for which evidence suggests the data aid public health professionals in notifiable disease surveillance processes. A number of peer-reviewed ELR studies (10, 16, 22, 28), as well as white papers published by public health professional organizations such as the International Society for Disease Surveillance (ISDS) and the Council on State and Territorial Epidemiologists (CSTE), discussed useful fields including sex, race, and ethnicity.

The final data set was augmented with "units of measure" as suggested by a group of experts working at and in close proximity to Indiana University's School of Medicine who were consulted for the project. The experts were provided with a draft data set that included the IAC and evidence-based elements. Some of the experts considered units to be helpful since many lab tests are often identical except for the kind of quantity examined in the specimen (29). For example, the concentration of sodium in a urine sample can be measured in terms of its mass concentration (ug/mL) or molar concentration (mmol/L).

Once the minimum data set was defined, the lead author mapped each data element to one or more corresponding fields from the HL7 Version 2 technical specification. The final data elements, their corresponding HL7 fields, and the source of their usefulness are summarized in Table 1.

Key Data Element	Corresponding HL7 Field(s)	Source
Patient's Identifier	Patient Identifier (PID-3)	(30)
Patient's Name	Patient Name (PID-5)	IAC
Patient's Date of Birth	Date of Birth (PID-7)	IAC
Sex (Gender)	Administrative Sex (PID-8)	(16, 28)
Race	Race (PID-10)	(28)
Patient's Address	Patient Address (PID-11)	IAC
Patient's Home Phone	Phone Number (PID-13)	(16, 28)
Number		
Ethnicity	Ethnic Group (PID-22)	(28)
Name of Attending Physician	Ordering Provider (OBR-16)	IAC
or Hospital or Clinic or	Ordering Facility Name (ORC-21)	
Submitter	Staff Name (STF-3)	
<b>Telephone Number of</b>	Order Callback Phone Number (OBR-	IAC
Attending Physician or	17)	
Hospital or Clinic or	Ordering Facility Phone Number (ORC-	
Submitter	23)	
	Staff Phone (STF-10)	
Address of Attending	Staff Office/Home Address (STF-11)	IAC
Physician or Hospital or Clinic	Ordering Provider Address (ORC-24)	
or Submitter		
Test Name	Observation Identifier (OBX-3)	IAC
<b>Test Results or Laboratory</b>	Observation Value (OBX-5)	IAC
<b>Interpretation of Test Results</b>		
Specimen Source	Specimen Source (OBR-15)	(16)
Units of Measure	Units (OBX-6)	Experts
Normal Range	Reference Range (OBX-7)	IAC
Abnormal Flag	Abnormal Flags (OBX-8)	(16)
Status of Test Result	Observation Result Status (OBX-11)	(22)

 Table 1 – A Minimum Data Set for Electronic Laboratory Reporting

 IAC = Indiana Administrative Code

Aims 2 and 3: Measurement and Comparison of Real-World ELR Data Completeness

The first sample contained 7,592,039 messages from the INPC's "raw" queue for incoming messages. In the raw sample, there were 7,592,039 possible values for fields within the PID segment, 7,471,001 possible values for fields within the OBR segment, and 22,244,305 possible values for fields within the OBX segment.

The second sample contained 16,365 messages from the Regenstrief NCD post-processed queue of reportable results. In the enhanced sample, there were 16,365 possible values within the PID segment, 35,266 possible values for fields within the OBR segment, and 131,665 possible values within the OBX segment.

Table 2 summarizes the calculated completeness for each field in the two samples. The first column contains the key data element name. The second column contains the corresponding HL7 field name. The third column contains the "percent complete" for each field in the raw sample. The fourth column contains the "percent complete" for each field in the raw sample. The fourth column contains the "percent complete" values across the samples.

Key Data Element	Corresponding HL7 Field	Percent Complete Raw	Percent Complete Enhanced	Difference
Patient's Identifier	Patient Identifier (PID-3)	99.9%	100%	+0.01%
Patient's Name	Patient Name (PID-5)	99.4%	100%	+0.06%
Patient's Date of Birth	Date of Birth (PID-7)	97.8%	99.8%	+2.0%
Sex (Gender)	Administrative Sex (PID-8)	95.8%	99.9%	+4.1%
Race	Race (PID-10)	38.4%	60.3%	+21.9%
Patient's Address	Patient Address (PID-11)	41.5%	63.3%	+21.8%
Patient's Home Phone Number	Phone Number (PID-13)	38.5%	72.8%	+34.3%
Ethnicity	Ethnic Group (PID-22)	3.5%	18.3%	+14.8%
Name of Attending Physician or Hospital or Clinic or Submitter	Ordering Provider (OBR-16)	57.4%	66.5%	+8.9%
Telephone Number of Attending Physician or Hospital or Clinic or Submitter	Callback Number (OBR-17) Staff Phone (STF-10)	0.15%	73.3%	+73.2%
Address of Attending Physician or Hospital or Clinic or Submitter	Staff Office/Home Address (STF-11)	N/A	84.6%	+84.6%
Test Name	Observation Identifier (OBX-3)	99.3%	100%	+0.07%
Test Results or Laboratory Interpretation of Test Results	Observation Value (OBX-5)	96.3%	98.9%	+2.6%
Specimen Source	Specimen Source (OBR-15)	13.7%	28.7%	+15.0%
Units of Measure	Units (OBX-6)	57.0%	17.5%	-39.5%
Normal Range	Reference Range (OBX-7)	55.8%	18.3%	-37.5%
Abnormal Flag	Abnormal Flags (OBX-8)	33.0%	28.4%	-4.6%
Status of Test Result	Observation Result Status (OBX-11)	92.8%	99.5%	+6.7%

 Table 2 – Comparison of INPC Completeness Profiles

The difference between fields across the two samples varied from 0.01% to 84.6%. The completeness for most of the fields increased, although several fields (Units of Measure, Normal Range, Abnormal Flag) decreased. The larger differences (Provider Phone Number, Provider Address) were observed for fields for which the data were directly enhanced by the INPC. Other fields varied in their completeness, although these variations are not attributable to the HIE's enhancement processes.

#### Discussion

To effectively perform surveillance and their other core functions, public health agencies require access to "timely, accurate, and complete data" (17). The results of this study confirm that laboratory data from clinical information

systems are heterogeneous in their completeness. In many cases, data important to public health surveillance processes are missing, indicating suboptimal ELR data quality. The study further demonstrates that HIEs employ methods that can mitigate ELR data deficiencies, improving the completeness of lab data electronically transmitted to public health information systems.

First, the study created a novel method for assessing the completeness of clinical data. While much of the literature on ELR and public health reporting focuses on improving the number of reportable cases submitted to public health agencies, this study measured the completeness of the data within individual reported cases. In previous studies, completeness of a data source is assigned a single value. For example, Effler et al. reported that the electronic communicable disease reporting system accounted for 91% of unique cases of notifiable disease (10). Heterogeneity of data completeness, however, makes it difficult to score an entire information system or data source as being 90% or 40% complete. A single score obscures whether the data source could adequately provide the data elements needed for recipient A versus recipient B. A system tracking spatial-temporal disease spread would benefit from a data source with a more complete address data. More complete address data would not, however, be useful for a statistical service that identifies when the number of reported cases rises above a certain threshold. Therefore this study assigned a percent complete to each of the data elements considered important to notifiable disease surveillance processes. A similar approach should be considered in the future when evaluating data sources to ensure that data consumers (humans or machines) understand the characteristics of the data from those sources.

The study also quantified what many in informatics are likely to encounter routinely: clinical data are heterogeneous in their completeness across and within information systems. Some laboratory information systems almost always transmit the specimen source (e.g., blood, urine), while others almost never provide this data element. Although this concept is not new, it is rarely measured and published.

Unfortunately many public health officials, like data consumers in other health care segments and industries, believe that data is easily and uniformly captured and stored across the spectrum of health care services. Data however are captured for a specific purpose, and the collection of additional data elements is costly. Additional data elements require staff to ask for and then record the information, which translates into additional time and labor. Therefore data consumers must understand the impact of the cost of data collection on the characteristics of data captured in various environments, like their completeness, when making decisions about secondary use. Public health officials, for example, might benefit from understanding that elements like the provider's phone number and address have little clinical relevance to the physician receiving the results of a lab test. These fields are poorly populated by laboratory information systems. Although these fields are required according to state (e.g., IAC) and federal (e.g., meaningful use) regulations, it does not guarantee that they will be complete and available for public health surveillance processes. Few addresses are provided today directly from the labs in the INPC; and very few phone numbers are provided. Thus policies to require additional data elements are unlikely to impact data collection processes.

Comparing the raw ELR messages with messages enhanced by the INPC demonstrates that the INPC employs methods that can improve the completeness of data. The completeness of nearly every field in the enhanced sample was larger than the equivalent fields in the raw sample. The improvements in completeness for provider names, addresses, and phone numbers were a direct result of HIE processes designed to enhance provider information. The INPC identifies all providers present anywhere in the message and resolves their identities using its Master Provider Index. The Master Provider Index is similar to master patient indices given that its function is to store a central list of all providers known to the INPC. The index possesses data elements such as the provider's name, clinic address, phone number, role (e.g., physician, physician assistant), and staff ID number. Using its Master Provider Index, the INPC is able to dramatically increase the amount of provider detail for the messages sent to the state health agency.

The Master Provider Index, however, was not specifically created for the public health reporting use case. The INPC has a more practical use of the index: the accurate delivery of lab results, radiology dictations, and other clinical documents to clinicians. There is intrinsic value to the INPC for knowing who providers are and where they practice to enable results delivery as well as other core HIE functions like access control. Such re-use of core HIE functionality is a benefit beyond improvements in data completeness. Leveraging core functions is one way that HIEs can support public health with little incremental cost. This is important, because in a recent survey of public health officials regarding participation in an HIE financial cost was a major concern (31). If multiple HIE participants are able to benefit from the same core set of technologies, then costs for all participants can be shared

and become more reasonable. Sustainability is a top priority for many HIEs, many of which struggle to support themselves when initial grant funding ends.

In addition to improving completeness and leveraging common infrastructure, HIE enhancements to laboratory data will improve public health surveillance and clinical workflow. Missing patient information necessitates a phone call from public health to request, for example, a patient's phone number. This would require the public health nurse to pause the investigation of a new notifiable disease case until the phone number could be identified. It would further require a nurse or other resource at the hospital or clinic to retrieve the voice mail, log into the EHR system (or pull a chart), extract the needed information, and call the public health department to provide the information. Inefficiencies due to data quality issues result in unnecessary costs and disruptions to routine clinical and public health workflows. Therefore any enhancements by HIEs to ELR data will improve public health surveillance processes for both clinicians and public health professionals.

While many fields in the enhanced messages have higher completeness, some fields have lower completeness. The differences between the raw and enhanced messages for these fields, however, are independent of the INPC's internal processes and enhancement methods. The INPC never removes data from a message; the HIE only adds information to or alters the value of a particular field. These differences reflect primarily heterogeneity in the data sources and message types. The raw sample consists of messages from 168 unique data senders, and the data pertain to all types of lab results (e.g., routine tests like white blood cell counts and hemoglobin A1c). The enhanced sample contains messages from 49 unique data senders and pertains only to positive notifiable disease results (e.g., sexually transmitted infections, lead levels in blood). A higher proportion of the enhanced sample contains microbiological cultures or micro results. For micro results, the units and normal range fields should be null as cultures are resolved through interpretation by a human lab technician. Future analyses of message completeness will control for this fact.

Abnormal result flags were also missing more often in the enhanced sample than the raw sample. Approximately five percent fewer abnormal flag values were observed. This outcome can also be explained by the fact that the enhanced sample contained a higher proportion of microbiological cultures. Micro results tend to be reported in a single field within the HL7 message (OBX-5), and some labs place micro results wholly in an NTE segment (a kind of comment field) at the end of the HL7 message. Values of "positive" embedded within an OBX-5 or comment field are challenging to process. Better use of abnormal flags would improve the Regenstrief NCD's ability, and other clinical information systems, to identify and route notifiable cases to clinicians and public health agencies.

#### Limitations

A limitation of this study is that the impacts upon public health surveillance processes are only estimated. While the literature provides some evidence on the impact of poor data quality, the specific impact of missing data in surveillance processes was not measured. Furthermore, measurable improvements to clinical and public health workflows as a result of INPC data enhancements were not captured in this study. This is work that researchers affiliated with the Indiana Center of Excellence in Public Health Informatics hope to perform in the future.

Additional work for the future includes the development and evaluation of processes to enhance patient-level data. The INPC plans to leverage its Master Patient Index in the same way that it currently leverages the provider index. We hypothesize that this will support a reduction in the number of calls to clinics and hospital wards to obtain additional details about patients who test positive for sexually transmitted infections and other notifiable diseases.

Finally the INPC is arguably one of the most robust and successful HIEs in the U.S. The INPC has partnered with local and state health agencies numerous times for over a decade to improve public health reporting, and the INPC has invested heavily in the development and maintenance of its Master Provider Index. Therefore the results of studies on data within in the INPC may not be generalizable to all HIEs and regions.

#### Conclusions

Poor quality data exists in clinical information systems, which presents a challenge for those interested in secondary uses of electronic health record data. For public health reporting, a single secondary use case, many data elements necessary to support surveillance processes are missing. Methods employed by HIEs to improve data quality can be leveraged by public health agencies to improve completeness, supporting both local needs to investigate disease outbreaks and federal goals to create meaningful use of EHR systems.

Although there is great opportunity for public health agencies, HIEs, hospitals, and laboratories to collectively improve public health surveillance processes, a number of challenges remain. Financial incentives to stimulate collaboration and data exchange may be necessary in some regions. Better use of existing standards, like the

abnormal flag field in HL7, will be necessary to improve identification of notifiable results. Finally, data consumer expectations need to be tempered to recognize not only the possibilities of HIE but also the limitations of certain data sources and systems.

Furthermore, the cost of collecting additional data in EHR and laboratory systems must be better understood by all stakeholders in health information exchange. Financial or other incentives may be required to drive changes to existing data collection workflow. New methods for collecting data *de novo* or leveraging existing data that minimize impact on workflow should be explored.

Ultimately, through research, development, and practice, we can build an information infrastructure capable of supporting secondary uses of electronic clinical data. This infrastructure will enable further improvements in public health surveillance processes not only in Indiana but across many states and regions.

#### Acknowledgements

The authors acknowledge Roland Gamache and Josette Jones, the other members of Dr. Dixon's doctoral research committee, for their guidance and support on the project and feedback on this paper. The authors further thank Shahid Khokhar of the Regenstrief Institute for his help with the extraction of the HL7 message data into a format suitable for analysis.

This study was supported, in part, by the Indiana Center of Excellence in Public Health Informatics through a grant award (1P01HK000077-01) from the U.S. Centers for Disease Control and Prevention. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

#### References

- 1. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007 Jan-Feb;14(1):1-9.
- 2. Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S, et al. Secondary use of clinical data. Studies in health technology and informatics. 2010;155:14-29.
- 3. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems. 1996;12(4):5-34.
- 4. Redman TC. The Impact of Poor Data Quality on the Typical Enterprise. Communications of the ACM. 1998;41(2):79-82.
- 5. Bailey JE, Pearson SW. Development of a tool for measuring and analyzing computer user satisfaction. Manaement Science. 1983;29(5):530-45.
- 6. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997 Sep-Oct;4(5):342-55.
- 7. Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. AMIA Annu Symp Proc. 2006:324-8.
- 8. Stein HD, Nadkarni P, Erdos J, Miller PL. Exploring the degree of concordance of coded and textual data in answering clinical queries from a clinical data repository. J Am Med Inform Assoc. 2000 Jan-Feb;7(1):42-54.
- 9. Doyle TJ, Glynn MK, Groseclose SL. Completeness of notifiable infectious disease reporting in the United States: an analytical literature review. Am J Epidemiol. 2002 May 1;155(9):866-74.
- 10. Effler P, Ching-Lee M, Bogard A, Ieong MC, Nekomoto T, Jernigan D. Statewide system of electronic notifiable disease reporting from clinical laboratories: comparing automated reporting with conventional methods. JAMA. 1999 Nov 17;282(19):1845-50.
- 11. AHRQ. HHS Awards \$139 Million To Drive Adoption of Health Information Technology. Rockville, MD: Agency for Healthcare Research and Quality; [updated October 13, 2004August 3, 2009]; Available from: http://www.ahrq.gov/news/press/pr2004/hhshitpr.htm.
- 12. Health Resources and Services Administration US. Justification of Estimates for Appropriations Committees. [updated 2007July 20, 2007]; Available from: ftp://ftp.hrsa.gov/about/budgetjustification08.pdf.

- 13. New York State Department of Health. Health Information Technology (Health IT) Grants HEAL NY Phase 1. 2006 [updated May; cited 2011 February 18]; Available from: http://www.health.state.ny.us/technology/awards/.
- 14. Robert Wood Johnson Foundation. Funding. 2010 [cited 2011 February 18]; Available from: http://www.projecthealthdesign.org/about/funding.
- 15. GAO. EMERGING INFECTIOUS DISEASES: Review of State and Federal Disease Surveillance Efforts. Washington, D.C.: U.S. Government Accountability Office2004. Report No.: GAO-04-877.
- 16. Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. Am J Public Health. 2008 Feb;98(2):344-50.
- 17. Lombardo JS, Buckeridge DL, editors. Disease Surveillance: A Public Health Informatics Approach. Hoboken: John Wiley & Sons; 2007.
- 18. Baxter R, Rubin R, Steinberg C, Carroll C, Shapiro J, Yang A. Assessing Core Capacity for Infectious Diseases Surveillance. The Lewin Group, Inc.; 2000 [cited 2010 March 8]; Available from: www.lewin.com/content/publications/808.pdf.
- 19. Lober WB, Karras BT, Wagner MM, Overhage JM, Davidson AJ, Fraser H, et al. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc. 2002 Mar-Apr;9(2):105-15.
- 20. AHRQ. AHRQ Research Relevant to Bioterrorism Preparedness. Rockville, MD: Agency for Healthcare Research and Quality; 2002 [updated March; cited 2010 March 17]; Available from: http://www.ahrq.gov/news/focus/bioterror.htm.
- 21. Magruder C. Public Health/Health Information Exchange Collaborative: A Model for Advancing Public Health Practice. Online Journal of Public Health Informatics [serial on the Internet]. 2010; 2(2): Available from: <a href="http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/ojphi/article/view/3217/0">http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/ojphi/article/view/3217/0</a>.
- 22. Panackal AA, M'Ikanatha N M, Tsui FC, McMahon J, Wagner MM, Dixon BW, et al. Automatic electronic laboratory-based reporting of notifiable infectious diseases at a large health system. Emerg Infect Dis. 2002 Jul;8(7):685-91.
- 23. Nguyen TQ, Thorpe L, Makki HA, Mostashari F. Benefits and barriers to electronic laboratory results reporting for notifiable diseases: the New York City Department of Health and Mental Hygiene experience. Am J Public Health. 2007 Apr;97 Suppl 1:S142-5.
- 24. Vreeman DJ, Stark M, Tomashefski GL, Phillips DR, Dexter PR. Embracing change in a health information exchange. AMIA Annu Symp Proc. 2008:768-72.
- 25. Zafar A, Dixon BE. Pulling back the covers: technical lessons of a real-world health information exchange. Studies in health technology and informatics. 2007;129(Pt 1):488-92.
- 26. Wurtz R, Cameron BJ. Electronic laboratory reporting for the infectious diseases physician and clinical microbiologist. Clin Infect Dis. 2005 Jun 1;40(11):1638-43.
- 27. Grannis SJ, Biondich PG, Mamlin BW, Wilson G, Jones L, Overhage JM. How disease surveillance systems can serve as practical building blocks for a health information infrastructure: the Indiana experience. AMIA Annu Symp Proc. 2005:286-90.
- 28. Vogt RL, Spittle R, Cronquist A, Patnaik JL. Evaluation of the timeliness and completeness of a Webbased notifiable disease reporting system by a local health department. J Public Health Manag Pract. 2006 Nov-Dec;12(6):540-4.
- 29. McDonald C, Huff SM, Mercer K, Hernandez JA, Vreeman DJ. Logical Observation Identifiers Names and Codes (LOINC®) Users' Guide. [PDF] Indianapolis: Regenstrief Institute; 2010 [cited 2011 January 28]; Available from: <u>http://loinc.org/downloads/files/LOINCManual.pdf</u>.
- 30. CSTE. Common Core Data Elements for Case Reporting and Laboratory Result Reporting. Atlanta: Council of State and Territorial Epidemiologists2009 Contract No.: 09-SI-01.
- 31. Hessler BJ, Soper P, Bondy J, Hanes P, Davidson A. Assessing the relationship between health information exchanges and public health agencies. J Public Health Manag Pract. 2009 Sep-Oct;15(5):416-24.

# Use of electronic health record data to evaluate overuse of cervical cancer screening

Jason S Mathias,<sup>1</sup> Dana Gossett,<sup>2</sup> David W Baker<sup>1</sup>

#### ABSTRACT

**Background** National organizations historically focused on increasing use of effective services are now attempting to identify and discourage use of low-value services. Electronic health records (EHRs) could be used to measure use of low-value services, but few studies have examined this. The aim of the study was to: (1) determine if EHR data can be used to identify women eligible for an extended Pap testing interval; (2) determine the proportion of these women who received a Pap test sooner than recommended; and (3) assess

the consequences of these low-value Pap tests. **Methods** Electronic query of EHR data identified women aged 30—65 years old who were at low-risk of cervical cancer and therefore eligible for an extended Pap testing interval of 3 years (as per professional society guidelines). Manual chart review assessed query accuracy. The use of low-value Pap tests (ie, those performed sooner than recommended) was measured, and adverse consequences of low-value Pap tests (ie, colposcopies performed as a result of low-value Pap tests) were identified.

**Results** Manual chart review confirmed query accuracy. Two-thirds (1120/1705) of low-risk women received a Pap test sooner than recommended, and 21 colposcopies were performed as a result of this lowvalue Pap testing.

**Conclusion** Secondary analysis of EHR data can accurately measure the use of low-value services such as Pap testing performed sooner than recommended in women at low risk of cervical cancer. Similar application of our methodology could facilitate efforts to simultaneously improve quality and decrease costs, maximizing value in the US healthcare system.

#### INTRODUCTION

Healthcare spending in the USA continues to increase more rapidly than inflation.<sup>1</sup> Most healthcare expenses are for services that improve patients' quality of life, longevity, or both. However, some tests are of low value-that is, they have marginal or no benefit, may harm patients, and waste financial resources. National organizations that have historically focused on increasing the use of effective services (eg, the American College of Physicians, the National Quality Forum, and the American Medical Association's Physician Consortium for Performance Improvement) are now attempting to decrease the use of low-value services.<sup>2-6</sup> To ultimately improve efficiency and reduce healthcare costs, methods will be needed to translate these recommendations into clinical practice.

Electronic health records (EHRs) have previously been utilized to increase the use of beneficial services, and they may similarly be utilized to decrease the use of low-value services.<sup>7 8</sup> However, experience with using EHRs to identify low-value services is limited. Secondary analysis of EHR data has been used to examine if life expectancy (and therefore relative value) was associated with variation in cancer screening practices among older patients.9 10 In a randomized controlled trial by Tierney and colleagues, computerized predictive information improved targeting of laboratory tests to higher-risk patients while decreasing use in lower-risk patients.<sup>11</sup> Bates and colleagues used EHR alerts to decrease the overuse of low-value, redundant laboratory tests.<sup>12</sup>

EHR data could similarly be used to identify lowvalue preventive services, and Pap tests are a prime target. Professional guidelines are in relative agreement that annual Pap tests are a low-value service for women at low risk of cervical cancer and therefore recommend an extended screening interval in these low-risk women.<sup>13–15</sup> However, annual Pap tests are a high-value service for women at high risk of cervical cancer, and professional guidelines recommend continued annual screening in high-risk women.<sup>13–15</sup> Therefore efforts to minimize the use of low-value Pap tests could result in harm if highrisk women were to erroneously receive an extended screening interval. The information necessary to determine cervical cancer risk, and therefore Pap test value, is uniformly reported in searchable EHR fields. However, it is unknown if determinations of Pap test value based on this EHR data are sufficiently precise for use in efforts to measure and decrease the use of low-value Pap tests.

The aims of the study were to: (1) determine whether it is possible to use EHR data to accurately identify women eligible for an extended cervical cancer screening interval; (2) use EHR data to determine the proportion of low-risk women eligible for an extended screening interval who received a Pap test sooner than recommended by current guidelines; and (3) determine the number of low-risk women who underwent a colposcopy as a consequence of a Pap test performed sooner than recommended by current guidelines.

#### METHODS

#### **Definition of low-value Pap tests**

Table 1 displays the cervical cancer screening guidelines at the time of the study.<sup>14–16</sup> If a woman eligible for triennial screening according to the guidelines received a Pap test sooner than recommended, that Pap test was considered to be a low-value Pap test.

► An additional appendix is published online only. To view this file please visit the journal online (www.jamai.org/content/ early/recent).

<sup>1</sup>Department of Internal Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA <sup>2</sup>Department of Obstetrics and Gynecology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

#### **Correspondence** to

Dr Jason S Mathias, Health Services Research Fellow, Institute for Healthcare Studies, Feinberg School of Medicine, Division of General Internal Medicine, 750 North Lakeshore Drive 10th Floor, Chicago, IL 60611, USA; j-mathias@md. northwestern.edu

An earlier version of the manuscript was presented as a poster at the Society of General Internal Medicine Annual Meeting in Minneapolis, MN in 2010 and as an oral presentation at the Academy Health Annual Research Meeting in Seattle, WA in 2011.

Received 10 August 2011 Accepted 17 December 2011

#### **Research and applications**

<b>Table 1</b> Cervical cancel screening practice guidelines current as of 2007	Table 1	Cervical cancer	screening	practice	guidelines	current as of 2007
---	---------	-----------------	-----------	----------	------------	--------------------

	American College of Obstetrics and Gynecology (2003)	American Cancer Society (2002)	US Preventive Services Task Force (2003)
Screening interval	For women $\geq$ 30, may extend interval to:	For women $\geq$ 30, may extend interval to:	At least every 3 years
	<ul> <li>Every 2–3 years if three consecutive tests are NIL</li> <li>Every 2 years if NIL test and negative HPV</li> </ul>	<ul> <li>Every 2–3 years if three consecutive tests are NIL</li> <li>Every 2 years if NIL test and pagetive HPV</li> </ul>	
	Every 3 years in INIL test and negative HPV	Every 5 years if NiL test and negative HFV	

HPV, human papilloma virus; NIL, negative for intraepithelial lesion.

#### **Practice setting and EHR**

We used patient data from the Northwestern Medical Faculty Foundation (NMFF) General Internal Medicine Clinic, an urban, academic, primary care practice with 38 general internal medicine attending and 51 resident physicians and approximately 60 000 clinic visits yearly. All physicians perform liquid-based Pap tests and use the Hybrid Capture II human papilloma virus (HPV) test. All physicians use an EHR for all clinical encounters (EpicCare; Epic Systems Corporation, Madison, Wisconsin, USA). The EHR has discrete fields for medical history, surgical history, current and past medications (including date of order and discontinuation), encounter diagnoses, and a problem list. Diagnosis names are linked to *International Classification of Diseases*, Ninth Revision (ICD-9) codes.

The EHR includes data from all specialties within NMFF and includes comprehensive clinical decision support for preventive care and disease management. The clinical decision support system includes a point of care reminder to perform Pap testing. The default interval is set at 1 year after the last Pap test result was recorded; providers can manually change the alert interval to biennial or triennial screening when indicated (ie, the provider determines that a woman is low risk). As of 2010, 31 of 38 (81.6%) general internal medicine attending providers had changed the Pap test alert frequency from 1 to 3 years for at least one of their patients.

#### **Eligibility criteria**

This study was approved by the institutional review board at Northwestern University. An electronic query identified all women 30-65 years old with one or more visits to any NMFF general internal medicine provider between January 1, 2007 and December 31, 2007. We defined low-risk women according to American Cancer Society, American College of Obstetrics and Gynecology, and US Preventative Services Task Force guidelines at the time of the study (table 1).<sup>14-16</sup> The electronic query identified women with a Pap test read as 'negative for intraepithelial lesion' (NIL) in 2007 and two prior NIL Pap tests in 2004-2006. These women were considered to be low risk and therefore eligible for an extended screening interval. In addition, women with a NIL Pap test and a negative HPV test in 2007 were considered to be low risk and therefore eligible for an extended screening interval (box 1). At the time of this study, manual chart review was necessary to determine HPV results. The medical history, problem list, encounter diagnoses, medication list, orders, and Pap test reports from January 1, 2004 to December 31, 2009 were queried to characterize these women, the care they received, and the results of testing. The electronic query can be found in the online appendix.

#### **Exclusion criteria**

We excluded women at high risk of cervical cancer because they are not eligible for extending the screening interval. High-risk women were identified via electronic query. As per professional guidelines, high risk was defined as a history of an abnormal Pap test between 2004 and 2007, cervical intraepithelial neoplasia (CIN) II or III, cervical cancer, HPV positive in 2007, diethylstilbestrol exposure in utero, immunosuppressive medication in 2007, or diagnosis associated with immunosuppression between 2004 and 2007 (box 1).<sup>14</sup> <sup>16</sup>

#### Box 1 Inclusion and exclusion criteria

#### Eligible for extended interval

Women 30-65 years old 2007 Pap test read as NIL

One of the following:

- Two NIL Pap tests between January 1, 2004 and December 31, 2006
- Negative HPV test in 2007

#### Ineligible for extended interval

Past medical history, encounter diagnosis, or problem list code for:

- Immunosuppression
- Solid organ transplant
- HIV
- Administration of chemotherapy
- Immune deficiency disorder
- Neutropenia
- Previously abnormal Pap tests
  - History of CIN II, III, or cervical cancer
  - Abnormal Pap test between 2004 and 2007
  - Colposcopy between 2004 and 2007
- Diethylstilbestrol exposure

Medication codes for:

- Azathioprine
- 6-Mercaptopurine
- Methotrexate
- Entanercept
- Tacrolimus
- Sirolimus
- Infliximab
- Adilimumab
- Muromonab-CD3
- Basiliximab
- Daclizumab
- Atgam
- Cyclophosphamide
- Cyclosporin
- Cyclospo
   Anakinra
- Mycophenolate mofetil
- Granulocyte colony-stimulating factor

CIN, cervical intraepithelial neoplasia; HPV, human papilloma virus; NIL, negative for intraepithelial lesion.

#### Analysis

To ensure patients were included appropriately, we manually reviewed the charts of 100 randomly selected women who the electronic query classified as low risk (ie, eligible for an extended screening interval). Data including age, race, marital status, number of visits per year, and number of chronic illnesses were extracted from the EHR database to characterize those women eligible for an extended screening interval.

Using EHR data, we identified all Pap tests performed during 2008 and 2009 in women who were eligible for an extended screening interval. These Pap tests were considered to be of low value because they were performed sooner than recommended. We identified all women in the cohort who had a colposcopy following the Pap test in 2008 or 2009. All colposcopies performed in these women were considered to have the potential to be adverse consequences of low-value Pap tests. We performed a separate manual review of the records belonging to those women receiving colposcopies, because we felt that these women were most likely to have been misclassified as low risk.

#### RESULTS

#### Low-risk women eligible for an extended screening interval

The EHR query identified 4002 women who had a NIL Pap test in 2007. Of these, 1749 were not eligible for an extended screening interval because they had not received two prior normal Pap tests during 2004–2006 or a negative HPV test in 2007. An additional 548 women were excluded after electronic query revealed an excluding diagnosis, excluding medication, or abnormal Pap test between 2004 and 2007. Ultimately, 1705 women were identified as eligible for an extended screening interval (figure 1). Patient characteristics are shown in table 2; 53.1% of eligible women were married, and 92.5% had  $\leq 1$ chronic condition.

#### **Confirming query accuracy**

We found 99 of the 100 women randomly selected for manual chart review were appropriately classified as eligible for an extended screening interval (ie, low risk). The one high-risk woman misclassified as low risk had a history of CIN of unknown severity that was recorded only in free-text notes.

#### Use of low-value Pap testing

Of the 1705 women identified as eligible for an extended screening interval by electronic query, 1120 (65.7%) received a low-value Pap test in 2008 or 2009 (figure 1). In 2008, 839 low-value Pap tests were performed, and 712 in 2009. A total of 431 (25.3%) women received a low-value Pap test in both years.

#### **Consequences of Pap testing sooner than recommended**

The electronic query identified 23 women who underwent colposcopy. Review of providers' electronic notes (of these 23 women) revealed that two women were misclassified as low risk (ie, had indications for annual Pap testing). One had a history of CIN II, and the other had a history of CIN of unknown severity; this information was recorded only in free-text notes. After exclusion of these two women, a total of 21 women (1.2% of those eligible for an extended screening interval) had undergone a colposcopy as a consequence of a low-value Pap test in 2008 or 2009 (figure 1). Five women who underwent colposcopy as a consequence of low-value Pap tests had CIN I. The remaining women had normal colposcopic findings or koilocytotic atypia consistent with HPV. Despite the lack of indication for an annual Pap, one woman's colposcopy led to a diagnosis of adenocarcinoma in situ (AIS).

#### Results using an extended screening interval of 2 years

If providers were following an extended screening interval of 2 years (the shortest interval recommended by the American College of Obstetrics and Gynecology and American Cancer

Figure 1 Identification of eligible women, low-value Pap testing, and its consequences. \*These two women were excluded after manual chart review of all patients receiving colposcopy as a consequence of lowvalue Pap testing. One had a history of cervical intraepithelial neoplasia (CIN) of unknown severity recorded only in provider notes. The other had a history of CIN II recorded only in provider notes. NIL, negative for intraepithelial lesion.



#### **Research and applications**

Table 2	Characteristics of women eligible for an
extended	screening interval (n=1705)

Characteristic	n (%)
Marital status	
Single	669 (39.2)
Married	904 (53.1)
Divorced	99 (5.8)
Widowed	26 (1.5)
Separated	5 (0.3)
Unknown	2 (0.1)
Race	
White	856 (50.2)
African-American	379 (22.2)
Latino	101 (5.9)
Other	42 (2.5)
Unknown	327 (19.1)
Number of GIM visits in 2007	
≤2	1190 (69.8)
3—5	416 (24.4)
≥6	99 (5.8)
Number of chronic conditions	
0	1146 (67.2)
1—2	535 (31.4)
≥3	24 (1.4)
Chronic conditions by disease	
Hypertension	367 (21.5)
Diabetes	79 (4.6)
History of cancer	9 (0.5)
Chronic pulmonary disease	222 (13.2)
Coronary artery disease	18 (1.1)
Congestive heart failure	5 (0.3)
Peripheral vascular disease	6 (0.4)
Chronic liver disease	3 (0.2)
Chronic kidney disease	4 (0.2)
Dementia	3 (0.2)

GIM, general internal medicine.

Society guidelines) instead of 3 years, then 839 (49.2%) women received a low-value Pap test in 2008. A total of 10 (0.6%) women received a colposcopy as a consequence of a low-value Pap test in 2008.

#### DISCUSSION

Clinical guidelines for preventive services such as Pap testing must always balance the benefits, harms, and costs of screening. Nationally representative models have determined that the risk of cervical cancer attributable to extended interval screening of low-risk women is approximately three per 100000 women.<sup>17</sup> The identification of one additional case of invasive cervical cancer requires a large number of additional tests at a significant cost.<sup>17</sup><sup>18</sup> On the basis of this information, professional societies have decided that the cost and harm of more frequent screening was not justified, and they recommended that low-risk women undergo triennial screening and avoid more frequent, low-value Pap testing.<sup>13–15</sup> <sup>19</sup> However, the guidelines maintain that annual Pap testing is of high value for women at high risk of cervical cancer.  $^{13-15}$  <sup>19</sup> Therefore Pap tests are a good example of a high frequency service with variable value that providers and policymakers should include in efforts to improve healthcare efficiency. while simultaneously maintaining efforts to ensure that all women receive cervical cancer screening when appropriate.

Our study did not intend to challenge or validate national cervical cancer screening recommendations. Rather, we sought

only to demonstrate how EHR data can be used to measure guideline-concordant care. However, our findings do bring the balance of risks and benefits of more frequent screening into sharp focus. A total of 21 women had negative colposcopies and likely experienced physical discomfort and psychological distress over possibly having cancer as a consequence of low-value Pap testing.<sup>20 21</sup> On the other hand, we did identify a woman with adenocarcinoma in situ (AIS) who might have been harmed if she had not been screened for an additional year; this is a predictable consequence of less frequent screening.

In our study, 66% of low-risk women received Pap tests sooner than recommended, representing inefficient care. These low-value Pap tests had significant downstream consequences. The unavoidable false-positive results and follow-up colposcopies associated with overscreening can cause undue psychological stress for women without cervical cancer.<sup>20</sup> <sup>21</sup> In addition, the financial burden of overscreening is substantial. Assuming a Pap test cost of US\$63 and a colposcopy cost of US\$286, the cost of low-value Pap tests and their consequences in our practice was approximately US\$100 000.<sup>22</sup> Although the results of this single-practice study cannot be extrapolated to the entire population, it is likely that low-value Pap tests cost the US healthcare system approximately US\$0.5–1 billion per year while achieving little or no improvements in health.<sup>22</sup> <sup>23</sup>

In order to eliminate this low-value spending, we must first have a reliable method to measure the use of low-value services. Our study successfully demonstrated that EHR data can be used for this purpose. The query accurately identified women at low risk of cervical cancer (and therefore low-value Pap testing), but it was not perfect. One percent of high-risk women were misclassified as low risk because information pertinent to assessment of cervical cancer risk was recorded only within freetext provider notes. Ideally, the query would be 100% specific to ensure that no harms could result from erroneously labeling a high-risk woman as low risk and advising an extended interval of 3 years. However, even if the specificity of queries can be improved, it is likely that there will always be a low rate of incorrectly labeling some women as low risk, which has potential harm.

This study had several limitations. First, we did not address all situations in which cervical cancer screening is of low value. National surveys have identified propensity for cervical cancer overscreening with respect to initiation of screening, screening in women after hysterectomies, failure to extend the screening interval when using a liquid-based Pap test, failure to stop screening in older women, and low-value use of the HPV test.<sup>24-28</sup> All could probably be identified with similar electronic queries and would be worthwhile investigating in the future. Second, we did not have results of HPV tests before 2007. If a woman had three NIL Pap tests and a positive HPV test in 2005 or 2006, she may have been misclassified as low risk. However, this is unlikely to significantly affect the results, as only a small minority of women seen received HPV testing at the time of the study. Third, by only reviewing Pap tests between 2004 and 2007, we may have inadvertently excluded those low-risk women not receiving Pap tests sooner than recommended. However, only 21 women had a NIL Pap test in 2004 with no subsequent Pap tests until 2007. Including this small number of low-risk women who are (presumably) receiving extended interval screening as recommended would not have significantly affected our results. Finally, the study was conducted in the general internal medicine practice of a single multispecialty group using a single EHR. Accuracy of the query may vary in other EHRs and at other practices.

Of note, the study practice's EHR includes a health maintenance prompt to perform Pap testing, and our practice has set the default frequency at 1 year (ie, the reminder occurs annually unless a clinician changes the reminder interval). This was done to ensure that high-risk women were not inadvertently screened every 3 years (ie, prevent underuse); this may have promoted overscreening in low-risk women. However, the rate of lowvalue Pap test use in our study was similar to that reported in national surveys, in which 65% of providers report performing Pap tests sooner than recommended.<sup>25</sup> Therefore we believe the default setting of 1 year used for the cervical cancer screening alert in our practice was not a major factor contributing to the high rate of low-value Pap tests performed in this study.

Nonetheless, it is important to acknowledge the ramifications of regularly using default settings for clinical reminder systems. Because default settings assume that a service has uniform risk and benefit across patients, the choice of a default setting can lead to overscreening (if the default setting is too short) or underscreening (if the default setting is too long). Algorithms such as ours can use EHR data to determine which setting is most appropriate for each individual patient. Such methods could be used to optimize preventive service use (ie, minimize both over- and under-use) and should be explored further as a means of ensuring the provision of necessary services while simultaneously decreasing costs.

Ultimately, the fundamental problem with increasing efficiency to cut healthcare costs is the inherent tension between wanting to make sure that patients receive all necessary tests (avoiding underuse) while simultaneously not using low-value services such as annual Pap testing in women at low risk of cervical cancer (avoiding overuse). To date, the problem of overuse has largely been ignored.<sup>29</sup> However, increasing health-care costs and pressure from health policymakers have spurred initiatives by several professional organizations to identify low-value services and recommend against their use.<sup>2–6</sup>

Ours and other studies suggest that these recommendations alone are insufficient to significantly reduce low-value service use in clinical practice.<sup>25</sup> We have demonstrated a method to leverage EHR data to guide interventions that could systematically reduce the use of low-value services identified by these professional societies while simultaneously maintaining efforts to reduce underuse and ensure that all patients receive necessary services.

However, any attempt to use EHR data in this manner is likely to be controversial, and steps should be taken to minimize potential harms. First, what constitutes a low-value service must be clearly defined and openly communicated to patients to ensure they do not feel necessary care is being withheld. Second, effective methods for preventing low frequency misclassification errors should be explored. Improved methods of data capture, such as natural language processing, may ultimately improve accuracy by identifying information buried in text fields. Even if such improvement can be achieved, real-time provider oversight is necessary to catch misclassification errors as they occur. For example, a provider reviewing a patient's newest Pap test results could be prompted to confirm the electronic query's determination that a woman has been newly determined to be at low risk of cervical cancer. Finally, interventions to decrease lowvalue service use should be implemented in conjunction with additional quality improvement efforts to ensure that the delivery of recommended preventive services continues.

#### CONCLUSIONS

We have shown that EHR data can be used to accurately identify low-value Pap test use, that low-value Pap testing occurs commonly, and that it has significant consequences. Similar application of our methodology could leverage increasingly available EHR data to systematically reduce low-value service use. However, the ramifications of using our methods in clinical practice are not well understood. Initial implementation should be coupled with research to identify optimal methods to address patient concerns, to minimize harms from misclassification errors, and to ensure continued delivery of recommended preventive services. Although efforts should proceed cautiously at first, our methodology could ultimately guide efforts to decrease low-value service use and maximize value in healthcare, simultaneously improving quality and decreasing costs in the US healthcare system.

 ${\bf Funding}~{\rm JSM}{\rm 's}$  fellowship is funded by AHRQ grant 5T32HS000078-13. The project is not supported in any other way.

#### Competing interests None.

Ethics approval IRB of Northwestern University.

**Contributors** All authors made substantial contributions to the conception and design of this work, acquisition and interpretation of the data, and writing of the manuscript. All authors reviewed the final version of the manuscript as submitted and approve it for publication. All those who are qualified to be authors are listed in the byline.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Study protocol and limited data available on request from Dr Jason S Mathias (email, jasonmathi@gmail.com).

#### REFERENCES

- Kaiser Family Foundation. Percent Annual Increase in National Health Expenditures (NHE) per Capita vs. Increase in Consumer Price Index (CPI), 1980-2009. 2011. http://facts.kff.org/chart.aspx?ch=212 (accessed 20 Jun 2011).
- American College of Physicians. ACP's High-Value, Cost-Conscious Care Initiative. 2011. http://www.acponline.org/advocacy/events/state\_of\_healthcare/ cost\_initiative11.pdf (accessed 10 Jun 2011).
- National Quality Forum. Overuse. 2011. http://www.qualityforum.org/Topics/ Overuse.aspx (accessed 20 Jun 2011).
- American Medical Association. PCPI and PCPI Approved Quality Measures. 2011. http://www.ama-assn.org/apps/listserv/x-check/qmeasure.cgi?submit=PCPI (accessed 24 Jun 2011).
- Chung J, Kaleba E, Wozniak G. Empirical Applications for Assessing the Efficiency and Value of Healthcare Using the Physician Consortium for Performance Improvement Physician Performance Measures. American Medical Association, 2008. http://www.ama-assn.org/ama1/pub/upload/mm/370/empirical\_applications.pdf (accessed 24 Jun 2011).
- Owens DK, Qaseem A, Chou R, et al. High-value, cost-conscious health care: concepts for clinicians to evaluate the benefits, harms, and costs of medical interventions. Ann Intern Med 2011;154:174–80.
- Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005;293:1223–38.
- Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006;144:742–52.
- Walter LC, Bertenthal D, Lindquist K, et al. PSA screening among elderly men with limited life expectancies. JAMA 2006;296:2336-42.
- Walter LC, Lindquist K, Nugent S, et al. Impact of age and comorbidity on colorectal cancer screening among older veterans. Ann Intern Med 2009;150:465-73
- 11. **Tierney WM**, McDonald CJ, Hui SL, *et al.* Computer predictions of abnormal test results. Effects on outpatient testing. *JAMA* 1988;**259**:1194–8.
- Bates DW, Kuperman GJ, Rittenberg E, et al. A randomized trial of a computerbased intervention to reduce utilization of redundant laboratory tests. Am J Med 1999;106:144–50.
- ACOG Committee on Practice Bulletins—Gynecology. ACOG Practice Bulletin no. 109: Cervical cytology screening. *Obstet Gynecol* 2009;114:1409–20.
- Saslow D, Runowicz CD, Solomon D, et al. American Cancer Society guideline for the early detection of cervical neoplasia and cancer. CA Cancer J Clin 2002;52:342-62.
- U.S. Preventive Services Task Force. Screening for Cervical Cancer: Recommendations and Rationale. 2003. http://www.uspreventiveservicestaskforce. org/3rduspstf/cervcan/cervcanrr.pdf (accessed 11 May 2011).
- ACOG Committee on Practice Bulletins. ACOG Practice Bulletin: clinical management guidelines for obstetrician-gynecologists. Number 45, August 2003. Cervical cytology screening (replaces committee opinion 152, March 1995). *Obstet Gynecol* 2003;102:417–27.

#### **Research and applications**

- Sawaya GF, McConnell KJ, Kulasingam SL, et al. Risk of cervical cancer associated with extending the interval between cervical-cancer screenings. N Engl J Med 2003;349:1501–9.
- Kulasingam SL, Myers ER, Lawson HW, et al. Cost-effectiveness of extending cervical cancer screening intervals among women with prior normal pap tests. Obstet Gynecol 2006;107:321-8.
- Wright TC Jr, Massad LS, Dunton CJ, et al. 2006 consensus guidelines for the management of women with abnormal cervical cancer screening tests. Am J Obstet Gynecol 2007;197:346-55.
- Wardle J, Pope R. The psychological costs of screening for cancer. J Psychosom Res 1992;36:609-24.
- Bell S, Porter M, Kitchener H, et al. Psychological response to cervical screening. Prev Med 1995;24:610-16.
- Saraiya M, McCaig LF, Ekwueme DU. Ambulatory care visits for Pap tests, abnormal Pap test results, and cervical cancer procedures in the United States. *Am J Manag Care* 2010;16:e137–44.
- Eltoum IA, Roberson J. Impact of HPV testing, HPV vaccine development, and changing screening frequency on national Pap test volume: projections from the National Health Interview Survey (NHIS). *Cancer* 2007;111:34–40.

- Meissner HI, Tiro JA, Yabroff KR, et al. Too much of a good thing? Physician practices and patient willingness for less frequent pap test screening intervals. *Med Care* 2010;48:249–59.
- Yabroff KR, Saraiya M, Meissner HI, et al. Specialty differences in primary care physician reports of papanicolaou test screening practices: a national survey, 2006 to 2007. Ann Intern Med 2009;151:602–11.
- Saraiya M, Berkowitz Z, Yabroff KR, et al. Cervical cancer screening with both human papillomavirus and Papanicolaou testing vs Papanicolaou testing alone: what screening intervals are physicians recommending? Arch Intern Med 2010;170:977–85.
- Holland-Barkis P, Forjuoh SN, Couchman GR, et al. Primary care physicians' awareness and adherence to cervical cancer screening guidelines in Texas. Prev Med 2006;42:140-5.
- Lee JW, Berkowitz Z, Saraiya M. Low-risk human papillomavirus testing and other nonrecommended human papillomavirus testing practices among u.s. Health care providers. *Obstet Gynecol* 2011;118:4–13.
- Health Care Efficiency Measures: Identification, Categorization, and Evaluation. Rockville, MD: Agency for Health Care Research and Quality, 2008. http://www.ahrq. gov/qual/efficiency/ (accessed 11 May 2011).

# Validity of electronic health record-derived quality measurement for performance monitoring

Amanda Parsons, Colleen McCullough, Jason Wang, Sarah Shih

#### ABSTRACT

Primary Care Information Project, New York City Department of Health and Mental Hygiene, New York, New York, USA

#### **Correspondence to**

Dr Amanda Parsons, Health Care Access & Improvement, New York City Department of Health and Mental Hygiene, 42-09 28th Street, 12th Floor, Queens, NY 11101, USA; aparsons@health.nyc.gov

Received 18 August 2011 Accepted 21 December 2011

Background Since 2007. New York City's primary care information project has assisted over 3000 providers to adopt and use a prevention-oriented electronic health record (EHR). Participating practices were taught to readjust their workflows to use the EHR built-in population health monitoring tools, including automated guality measures, patient registries and a clinical decision support system. Practices received a comprehensive suite of technical assistance, which included quality improvement, EHR customization and configuration, privacy and security training, and revenue cycle optimization. These services were aimed at helping providers understand how to use their EHR to track and improve the quality of care delivered to patients. Materials and Methods Retrospective electronic chart reviews of 4081 patient records across 57 practices were analyzed to determine the validity of EHR-derived quality measures and documented preventive services. **Results** Results from this study show that workflow and documentation habits have a profound impact on EHR-derived guality measures. Compared with the manual review of electronic charts, EHR-derived measures can undercount practice performance, with a disproportionately negative impact on the number of patients captured as receiving a clinical preventive service or meeting a recommended treatment goal. Conclusion This study provides a cautionary note in using EHR-derived measurement for public reporting of provider performance or use for payment.

Unlike most claims-based quality measurement, measures derived from EHR can incorporate clinical

findings, allowing for the tracking of intermediate outcomes such as blood pressure and body mass index. However, documentation habits by providers can vary, and the necessary data entered into the EHR may not be interpreted or recognized by standard EHR software programming. This may lead to undercounting the patients eligible for a preventive service (eg, diagnosis of ischemic cardiovascular disease) or receiving a recommended treatment (eg, screening or medication) or meeting a recommended target (eg, control of blood pressure to less than 140/90 mm Hg).

Formed in 2005, the New York City Primary Care Information Project (PCIP) has assisted over 3000 providers to adopt and use a prevention-oriented EHR as a means to improve the delivery of primary care. Nearly 40% of the participating providers are operating in small (fewer than 10 providers), physician-owned practices. PCIP selected an EHR vendor through a competitive process and codeveloped prevention-oriented functionality and population health monitoring tools, including automated quality reporting and a clinical decision support system (CDSS). The quality reporting tool displays by measure, for each eligible patient, whether the practice has or has not met the recommended preventive service. In addition, at the point of care, the CDSS function displays the preventive services a patient is eligible for and has not yet received, allowing the provider to take action (eg, order a mammogram, adjust medications, discuss smoking cessation aids) during the visit.

Providers were trained by both the EHR vendor's training staff and practice consultants employed by PCIP, who provided on-site technical assistance. Providers were taught to re-adjust the practice's workflows to document diagnoses and key preventive services in structured fields that are searchable and capable of generating the quality measures and preventive service reminders. Providers were also shown how to view their EHR calculated quality measures both within the EHR and through monthly reports generated by PCIP staff and emailed to individual providers. In addition, efforts were made to create alignment between payment and improved quality of care by informing providers of the various available incentive programmes as well as launching PCIP designed programmes. Through these synergistic changes: (1) prevention-oriented EHR; (2) practice workflow redesign; and (3) payment that rewards prevention, PCIP worked with primary care providers to prioritize prevention and facilitate management of chronic disease.<sup>3</sup>

This report provides an assessment of the validity of quality measures derived from

The American Recovery and Reinvestment Act of 2009 authorized US\$19 billion in funding for the deployment and meaningful use of electronic health records (EHR), and introduced a national framework for the adoption of health information technology.<sup>1</sup> The Center for Medicare and Medicaid Services has offered eligible providers financial incentives for demonstrating meaningful use of EHR and reporting on the quality of care.<sup>2</sup> Starting with stage 1 of meaningful use, the Center for Medicare and Medicaid Services calls for the submission of provider-level quality measures, initially by attestation but then through electronic submission, starting as early as 2012. Many stakeholders, including payers, independent physician associations, and consumers have a vested interest in accessing and utilizing EHR-derived quality measures for purposes of accountability or rankings. However, quality measures derived from EHR have yet to be validated as representative of provider performance for incentives or comparative purposes.

#### **Research and applications**

information entered into the EHR and describes the issues contributing to variations in the results of automated EHR quality measurement.

#### METHODS

#### **Practice selection**

A subset of 82 practices enrolled in a pilot rewards and recognition programme were invited to participate in the data validation study, as they had all implemented the eClinicalWorks EHR software before January 2009 and received technical assistance through the PCIP programme, and had a majority of their patient panel recorded in the EHR. Practices were required to have a minimum of 200 electronic patient records with a diagnosis of diabetes, hypertension, dyslipidemia, or ischemic cardiovascular disease. All 82 practices received a software upgrade between February and August 2009 to implement automated quality measurement reporting and the CDSS functionality. Participating practices signed a letter of consent allowing independent medical reviewers to conduct abstraction of the EHR and received an honorarium of US\$500 for completing the study. This study was approved by the Department of Health and Mental Hygiene institutional review board no 09-067.

#### **Electronic chart reviews**

Medical reviewers randomly sampled 120 electronic patient charts per practice for established patients between 18 and 75 years of age, with at least one office visit since the practice implemented the EHR. For this study, data from the manual review of the electronic chart (e-chart) were analyzed if the patient had an office visit during the 6-month period after the activation date of the quality measurement reporting tool and implementation of the CDSS.

For each e-chart, reviewers abstracted the patient's age and gender, along with vitals, diagnoses, medications, laboratory results, diagnostic images, vaccinations, and receipt of or referral to counselling for the most recent visit. Depending on the data element, reviewers were instructed to search in multiple locations of the EHR: problem list, medical history, social history, progress notes (chief complaint, history of present illness, assessment), procedures, diagnostic images, vitals, and laboratory tests. Each reviewer was trained and tested by a standardized approach to ensure interrater reliability. If there was uncertainty whether documentation would meet the quality measure criteria a senior reviewer and PCIP staff would determine whether to include or exclude the observation.

#### **Analytical methods**

For each patient, each data element was coded based on whether it was documented in a structured field recognized by the existing EHR software for automated quality measurement (1=location recognized and 0=not recognized for quality measurement reporting). In addition, two sets of numerator and denominator counts were generated for each of the 11 quality measures. The first set of counts included only those patients whose information was documented in structured fields recognized by the existing software (EHR automated). The second set of counts incorporated all information about patients documented in the EHR (e-chart review). Data element coding and all counts were calculated using Microsoft Access structured query language.

Simple frequencies and descriptive statistics were generated to determine data element documentation patterns and estimate population-level numerators and denominators for each quality measure. Wilcoxon non-parametric tests were used to compare practice-level numerators and denominators calculated for each measure.

All descriptive statistics and tests were conducted using SAS V.9.2 analytical software. Practice distribution of documentation and calculated proportions of elements documented in various locations of the electronic chart were generated in MS Excel and MS Access.

#### RESULTS

Of the 82 practices eligible for the study, 57 practices agreed to allow PCIP to conduct the e-chart review. Based on self-reported information and practice registries, practices that agreed to participate in the study had a higher average percentage of Medicaid insured patients (42.7% vs 30.7%), and a higher number of patients with diabetes (209.9 vs 110.0) and hypertension (477.8 vs 353.2) (table 1).

A total of 4081 e-charts was available for this analysis. An additional 2759 e-charts were reviewed, but were excluded from the analysis because the patients did not have a qualifying office visit during the 6-month study period. More than half of the final study sample of patients were women (59.4%), and the average patient age was 48.1 years. Participating practices varied in their distribution of patients with diagnoses of diabetes, ischemic cardiovascular disease, hypertension, dyslipidemia and patients who were current smokers (table 1). The majority (89.9%) of participating providers were primary care providers (ie, internal medicine, family medicine, pediatrics, obstetrics/gynecology). Non-primary care providers specialized in cardiology, pulmonology, endocrinology, allergy, gastroenterology, or did not specify a specialty (data not shown).

We looked across the 11 clinical quality measures to assess where information was documented. The presence of data recognized for automated quality measurement varied widely, ranging from 10.7% to 99.9% (table 2). Measure components

<b>Table 1</b> Unaracteristics of practices and patient charts revie
--

	Participating in chart review (N=57) Mean (min-max)	Not participating in chart review (N=28) Mean (min-max)
Practice characteristics		
No of providers per practice	2.9 (1-24)	2.4 (1-13)
No of full-time equivalent	1.7 (0.7-9.2)	1.9 (1.0-10.8)
Percentage of patients with Medicaid insurance or uninsured	42.7 (0.0-90.0)	30.7 (2.0-81.0)
Estimated population*		
Patients >18 years of age with at least one office visit in the past year	1936.2 (0-9117)	1905.5 (12-9723)
Diabetes	209.9 (3-1480)	110.0 (1-450)
Hypertension	477.8 (2-2460)	353.2 (1-1121)
Months using EHR by 1 July 2009	14.3 (7.5-39.1)	13.7 (3.7-22.4)
	Mean (Min – Max)	Not applicable
Patient characteristics of charts reviewed		
Medical charts reviewed (total 4081)	71.6 (19-111)	_
%Female	59.4 (35.0-80.7)	-
Patient age, years	48.1 (32.8-61.5)	-
Percentage of records with diagnoses		
Diabetes	17.6 (0.0-48.1)	_
Ischemic cardiovascular disease	7.3 (0.0-39.0)	_
Hypertension	41.2 (8.2-87.0)	_
Dyslipidemia	36.0 (3.3-85.4)	_
Current smokers	10.3 (1.4-37.9)	_

\*From automated reporting from the electronic health record (EHR)—separate from electronic chart reviews.

#### **Research and applications**

#### Table 2 Frequency of clinical information and locations in electronic charts for quality measurement

		scription of denominator or numerator (N) data element	Recognized for quality measurement			Not recognized for quality measurement		
Name of measure and brief description	(no	of eligible charts)	Location	No	%	Location	No	%
Screening measures (patients for the denom	inator	are identified by their age and gender)						
Breast cancer screening								
Female patients $\geq$ 40 years of age	Ν	Diagnostic order and result for	Procedures	29	10.7	Scanned patient docs	270	48.7
who received a mammogram in the		mammogram (543)				Diagnostic imaging	204	36.8
past 2 years						Other	10	3.6
BMI								
Patients $\geq$ 18 years of age who have	Ν	BMI documented (3122)	Vitals	3116	99.8	Medical history	2	0.1
a BMI measured in the past 2 years						Other	4	0.1
Influenza vaccine								
Patients $>$ 50 years of age who	Ν	Influenza vaccination documented (480)	Immunizations	475	99.0	Chief complaint/HPI	3	0.6
received a flu shot in the past year						Procedures	2	0.4
Smoking status								
Smoking status updated annually	Ν	Smoking status documented (3357)	Smart form	1796	53.4	Social history	1530	45.5
in patients $\geq$ 18 years of age		-				Other	26	0.8
						Medical history	5	0.2
Intervention measures								
Antithrombotic therapy								
Patients $\geq$ 18 years of age with	D	Diabetes diagnosis and age	Problem list	623	91.6	Medical history	42	6.2
a diagnosis of IVD or $\geq$ 40 years		≥40 years (680)				Assessment	11	1.6
of age with a diagnosis of						Chief complaint/HPI	4	0.6
diabetes taking aspirin or	D	IVD diagnosis (303)	Problem list	250	82.5	Medical history	45	14.9
						Assessment	7	2.3
						Chief complaint/HPI	1	0.3
	N	Appropriate medication prescribed (759)	Medications	757	99.7	Plan/treatment	2	0.3
BP control			modiodiono	101	00.7		-	0.0
Patients 18–75 with a diagnosis of	П	Hypertension diagnosis (1676)	Problem list	1497	89.3	Medical history	157	9.4
hypertension, with or without IVD,	5			1107	00.0	Assessment	11	0.1
with a recorded systolic BP						Chief complaint/HPI	11	0.7
<140 mm Hg and diastolic BP	Con	norbid diagnoses				onici complant/in i		0.7
< 90 mm Hg in the past 12 months	п	Diabetes (551)	Problem list	512	92.9	Medical history	28	51
patients with hypertension and DM)	D	Diduetes (331)	Troblem list	512	52.5	Accessment	10	1.0
						LDI	10	1.0
	п	IVD (118)	Problem list	02	70 0	Modical history	24	20.2
	D		FIUDIEIII IISL	93	70.0	Accomment	24	20.3
	N	PD decumented (2060)	Vitala	2066	00.0	Assessment Medical biotomy	ו י	0.9
Chalasteral association and control		BF documented (3000)	Vildis Drahlam liat	3000	99.9 7E 1	Medical history	260	10.1
cholesteror screening and control	D	Dystipideinia diagnosis (1466)	FIODIEITI IISL	1112	75.1		209	10.Z
						Assessment	00	0.0
lish dala						Chief complaint/HPI	12	0.8
High risk	<b>C</b> • •							
of age with a diagnosis of	Con	Diskatas (515)	Dashlara Kat	477	02.0	Madiaal history	20	го
dyslipidemia and DM or IVD with	U	Diabetes (515)	Problem list	4//	92.0		30	5.8
a measured LDL in the past year						Assessment	6	1.2
Out of the Development of the state of the		N/D (011)	Durble of Part	100		Unier complaint/HPI	2	0.4
Lontrol: Screened patients with	D	IVD (211)	Problem list	182	86.3	Nedical history	27	12.8
						Assessment	1	0.5
						Chief complaint/HPI	I	0.5
General population				1001			1100	45.0
Screening: M 35+, F45+ W/no	N	LDL cholesterol test result (2425)	Laboratory tests	1294	53.4	Scanned patient docs	1106	45.6
cholesterol or IDI in the past						Uther	24	1.0
5 years						Nedical history	1	0.0
Control: Screened patients with	Ν	Total cholesterol test result (2545)	Laboratory tests	1383	53.4	Scanned patient docs	1137	44.7
total cholesterol <240 mg/dl or			-			Other	24	0.9
LDL <160 mg/dl						Medical history	1	0.0
Hemoglobin A1c screening and control						,		
Screening: Patients 18-75 years	D	Diabetes diagnosis (740)	Problem list	676	91.4	Medical history	48	6.5
of age with a diagnosis of DM with		- · ·				Assessment	12	1.6
a documented hemoglobin A1c test						Chief complaint/HPI	4	0.5
within the past 6 months	•							
Control: Screened patients with	N	Hemoglobin A1c test result (642)	Laboratory tests	294	63.0	Scanned patient docs	171	36.6
						Uther	2	0.4
							Con	tinued

Parsons A, McCullough C, Wang J, et al. J Am Med Inform Assoc (2012). doi:10.1136/amiajnl-2011-000557

Research ar	nd applic	ations
-------------	-----------	--------

Table 2 Continued	b
-------------------	---

Name of measure and brief description		scription of denominator or numerator (N) data element	Recognized for quality measurement			Not recognized for quality measurement		
		of eligible charts)	Location	No	%	Location	No	%
Smoking cessation intervention								
Current smokers who received	D	Documented current smokers (409)	Smart form	243	59.4	Social history	161	39.4
cessation interventions or counselling in the past 12 months					Other	3	0.7	
					Assessment	2	0.5	
	Ν	Smoking cessation intervention (129)	Smart form	85	64.9	Other	33	25.2
						Medications	8	6.1
						Scanned patient docs	3	2.3

BMI, body mass index; BP, blood pressure; DM, diabetes mellitus; HPI, history of present illness; IVD, ischemic cardiovascular disease; LDL, low-density lipoprotein.

relying on vitals, vaccinations, and medications had the highest proportion of information documented in structured fields recognized by the automated quality measures. The majority of diagnoses for chronic conditions such as diabetes (>91.4% across measures), hypertension (89.3%), ischemic cardiovascular disease (>78.8% across measures) and dyslipidemia (75.1%) were documented in the problem list, a structured field used for automated quality measurement. Patient diagnoses not recognized for inclusion in the measure were recorded in the medical history, assessment, chief complaint, or history of present illness, sections that typically allow for free-text entries.

Diagnostic orders or results for mammogram had the lowest proportion (10.7%) of data recorded in structured fields recognized for automated quality measurement. The majority of the information for breast cancer screening was found as scanned patient documents and diagnostic imaging; both sources of information are not amenable for automated electronic queries.

Nearly half of the information for measures that require a laboratory test result, such as control of hemoglobin A1c and cholesterol, was documented in structured fields recognized for automated quality measurement (range 53.4–63.0%). Similarly, only half of the information regarding patient smoking status (53.4%) was recognized for automated quality measurement. With the exception of medications, vaccinations, and blood pressure readings, practices varied substantially in where they chose to document the data elements required for automated quality measurement (figure 1).

In estimating the denominator loss due to unrecognizable documentation, the average practice missed half of the eligible patients for three of the 11 quality measures—hemoglobin A1c control, cholesterol control, and smoking cessation intervention (table 3). No statistically significant differences were observed between the e-chart and EHR automated quality measurement scores in the number of patients captured for the denominator for the remaining eight measures. Current EHR reporting would underreport practice numerators for six of the 11 measures hemoglobin A1c control, hemoglobin A1c screening, breast cancer screening, cholesterol control, cholesterol screening, and smoking status recorded.

#### DISCUSSION

As the nation continues to drive EHR adoption through significant infusions of funding for health information technology infrastructure and support, and payment reform carries the promise of improved quality at lower cost, it is important and timely to assess the validity of EHR-derived clinical quality measures.

Figure 1 Distribution of documentation and data elements recognized for quality measurement across 57 practices. EHR, electronic health record; IVD, ischemic cardiovascular disease; LDL, low-density lipoprotein.



Table 3	Comparison of numerator and denominator counts for EHR automated quality measurement and
electronic	chart review (n=57 practices)

	Denominator				Numerator			
	EHR query		e-Chart review		EHR query		e-Chart review	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A1c control	4.9	5.2	8.2**	6.1	1.8	2.4	2.7*	2.6
A1c screening	11.8	8.6	12.9	8.6	4.9	5.2	8.2**	6.1
Antithrombotic therapy	13.1	10.3	14.7	10.6	7.5	7.8	8.2	7.8
Body mass index	71.4	18.1	71.4	18.1	54.6	23.2	54.8	23.2
Blood pressure control	25.7	14.4	29.3	14.4	14.1	9.6	16.2	10.0
Breast cancer screening	28.7	9.9	28.7	9.9	1.0	2.3	9.5**	8.2
Cholesterol control	15.7	14.3	29.1**	15.7	13.0	12.5	23.8**	13.5
Cholesterol screening	41.7	15.2	44.2	15.8	15.7	14.3	29.1**	15.7
Influenza vaccination	34.4	15.1	34.4	15.1	8.1	8.7	8.3	8.9
Smoking cessation intervention	4.3	3.8	7.2**	4.8	1.5	2.7	2.3	3.5
Smoking status recorded	71.5	18.1	71.5	18.1	31.5	24.0	59.5**	18.9

\*p<0.01 \*\*p<0.001.

EHR, electronic health record.

Data from this study suggest that clinical diagnoses for the majority of patients are documented in structured fields needed for automated quality reporting and point-of-care reminders, and that there is not a significant amount of 'denominator loss' when using EHR-derived measurement versus e-chart reviews. However, EHR-derived quality measurement can result in significant numerator loss, resulting in underestimates of the delivery of clinical preventive services.

It is important to understand the impact of workflows and documentation habits on EHR-derived quality measures. In this study, the majority of practices correctly documented diagnoses of hypertension and diabetes over 80% of the time, but rates of appropriate documentation for dyslipidemia and ischemic cardiovascular disease were substantially lower. Providers may be more likely to overlook chronic diseases that are documented elsewhere in the chart, a finding we commonly see in the case of obesity and active smoking. Anecdotally, providers have reported to us that they limit the number of assessments assigned to a patient at any given visit due to the historical limitation related to paper claims (some payers limit the number of assessments that can be reported in paper claims to four or fewer), which may have the unintended consequence of 'underdocumentation'.

For laboratory tests, the presence or absence of laboratory interfaces and the appropriate logical observation identifiers names and codes creates significant variation in practice scores. Even if a practice has an electronic interface, reference laboratories do not consistently provide EHR vendors with compendiums that have logical observation identifiers names and codes codes for each test, and therefore, some test orders and results remain undetected by EHR software. For practices with no electronic laboratory interface, the quality measures that rely on laboratory data can only be addressed by manually entering the results into structured fields, which few providers do.

In addition, for quality measures that rely on tests or procedures performed in a different office setting, the difficulty of getting the results back, and in structured form, makes it challenging to satisfy those measures. In this study, the mammography quality measure was only satisfied by the presence of a structured test result, yet most mammography results come back in the form of faxed results, necessitating an additional step by the practice to codify the results in structured form. Eventually, developments in natural language processing may help convert unstructured text results into their structured counterparts needed to satisfy quality measures and trigger clinical decision support. Finally, some quality measures may be impacted by incorrect or imprecise logic used to code the measure. For instance, with mammography rates, the relatively low practice performance seen in this study is largely attributable to a specific flaw in the design of the EHR's quality measure. For over a year, this particular measure would only allow a mammogram result to satisfy the measure if the results were structured and the test was ordered as a 'procedure', and not a 'diagnostic image', the latter being the more consonant with provider preference.

Providers uniformly scored well on documentation of aspirin therapy, influenza vaccination and blood pressure, probably due to the relative lack of options in the EHR to document these data in fields other than the designated locations. For some other measures, such as smoking status, documentation can vary widely provider to provider, including a variety of notations (eg, '+ smoker', '+ cig', '2 ppd', 'smoker', and '+ tobacco') and locations (eg, smart form, preventive medicine, social history). This variability in documentation preference has been shown to lead to significant variations in quality measurement depending on which fields are chosen and how much granularity is provided. The bimodal distribution of practice documentation for smoking status is probably due to the difference in whether practices used the structured fields in the smart form to satisfy the quality measure.

This study has several limitations. Several practices refused to participate in the chart review. The majority that did not participate stated they lacked sufficient physical space for the chart reviewers to conduct their reviews. Practices electing not to participate in the study did not differ in their population characteristics, as shown in table 1. This study also limited its chart reviews to practices that were using eClinicalWorks; therefore, the findings may not be generalizable to other EHR. In addition, the study focused on available documentation in the electronic record, and we did not conduct an audit of whether providers or practice staff actually delivered the services recorded. Separate studies will need to be conducted to ascertain whether information recorded in the EHR may not reflect the actual care delivered.

Finally, this study did not assess how practice characteristics, interventions by PCIP, or the use of specific EHR functionality may have impacted differences in documentation variation. These studies are being conducted separately.

Until recently, quality measures were largely derived from manual paper chart reviews, secondary analysis of claims

#### **Research and applications**

databases, or patient surveys, as opposed to being calculated from EHR derived data. This was driven in large part by the historically low rates of EHR use nationwide and the ensuing lack of data availability, quality, and comparability. Studies comparing claims data with clinical data have noted significant disparities between the two sources,<sup>4 5</sup> yet claims-based quality measurement has continued to be the dominant form of largescale quality analysis because no other data sources have been readily available either as a complement or replacement. In some measures, such as breast cancer screening, the use of administrative or claims data may still be more reliable until health information exchanges are established, more broadly adopted, and can integrate multiple data sources to establish more comprehensive measures on recommended care delivery and health outcomes.

EHR offer new potential for performance measurement given that most commercially available EHR use standard dictionaries to capture information in coded forms, such as ICD for problem list and SNOMED for medications. Using these codified data, EHR can help identify patient populations and calculate a significant number of quality measures that leverage data available in the EHR.<sup>6</sup> These measures can range from adherence to clinical guidelines to assessments of rates of clinical preventive services to rates of screening. However, EHR-derived quality measurement has limitations due to several factors, most notably variations in EHR content, structure and data format, as well as local data capture and extraction procedures.<sup>7 8</sup>

Several steps can be taken to mitigate the variability of EHR documentation. As part of PCIP's programme, providers are trained on proper documentation techniques during their initial EHR training, and then quality improvement specialists reinforce their use, but ultimately there are no mechanisms to force providers to document in a particular location in the chart. Providers need regular prompts, training and feedback to alter their documentation habits. Studies have shown that clinical decision support can help improve the quality and accuracy of documentation.9 Another way to mitigate the variability of documentation would be to include claims data to populate the EHR, thereby providing a more robust and complete profile of the patient. In addition, standards need to be developed for what needs to be documented in the various medical record components, such as a clinical encounter note or a care plan document. Much work is being done to standardize the output of EHR for use in health information exchange (eg, the continuity of care document), but few efforts are aimed at standardizing what data inputs should go into the EHR.

More studies are needed to assess the validity of EHR-derived quality measures and to ascertain which measures are best calculated using claims or administrative data or a combination of data sources. If provider-specific quality measurements are to be reported and made public, as is the plan for the meaningful use quality measures, further analysis is needed to understand the limitations of these data, particularly if they are prone to underestimation of true provider performance.

**Acknowledgments** The authors would like to acknowledge Dr Farzad Mostashari for his initial concept and design of PCIP as well as his vision for the preventionoriented EHR tools deployed in this project. They would also like to thank the participating practices and the Island Peer Review Organization staff for their dedication and many hours spent conducting chart reviews. They also wish to thank the PCIP staff for their tireless dedication to improving health.

**Funding** This study was supported by the Agency for Healthcare Research and Quality (grant nos R18HS17059 and 17294). The funder played no role in the study design, in the collection, analysis and interpretation of data, in the writing of the report or in the decision to submit the paper for publication.

#### Competing interests None.

Ethics approval This study was approved by the Department of Health and Mental Hygiene institutional review board no 09067.

Provenance and peer review Not commissioned; externally peer reviewed.

#### REFERENCES

- Blumenthal D. Promoting use of health IT: why be a meaningful user? Md Med 2010;11:18–19.
- CMS Office of Public Affairs. Fact Sheet. Details for Medicare and Medicaid Health Information Technology: Title IV of the American Recovery and Reinvestment Act. 2009. http://www.cms.hhs.gov/apps/media/press/factsheet.asp?Counter=3466& intNumPerPage=10&checkDate=&checkKey=&srchType=1&numDays=3500& srchOpt=0&srchData=&keywordType=All&chkNewsType=&6intPage=& showAll=&pYear=&year=&desc=&cboOrder=date (accessed 18 Aug 2011).
- Frieden TR, Mostashari F. Health care as if health mattered. JAMA 2008;299:950-2.
- Jollis JG, Ancukiewicz M, DeLong ER, et al. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. Ann Intern Med 1993;119:844–50.
- Tang PC, Ralston M, Arrigotti MF, et al. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. J Am Med Inform Assoc 2007;14:10–15.
- Roski J, McClellan M. Measuring health care performance now, not tomorrow: essential steps to support effective health reform. *Health Aff (Millwood)* 2011;30:682–9.
- Kahn M, Ranade D. The impact of electronic medical records data sources on an adverse drug event quality measure. J Am Med Inform Assoc 2010;17:185–91.
- Chan K, Fowles J, Weiner J. Electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev* 2010;67:503–27.
- Galanter WL, Hier DB, Jao C, et al. Computerized physician order entry of medications and clinical decision support can improve problem list documentation compliance. Int J Med Inform 2010;79:332-8.

#### ORIGINAL REPORT

# The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction

Richard Platt<sup>1</sup>\*, Ryan M. Carnahan<sup>2</sup>, Jeffrey S. Brown<sup>1</sup>, Elizabeth Chrischilles<sup>2</sup>, Lesley H. Curtis<sup>3</sup>, Sean Hennessy<sup>4</sup>, Jennifer C. Nelson<sup>5</sup>, Judith A. Racoosin<sup>6</sup>, Melissa Robb<sup>6</sup>, Sebastian Schneeweiss<sup>7</sup>, Sengwee Toh<sup>1</sup> and Mark G. Weiner<sup>8</sup>

<sup>1</sup>Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

<sup>2</sup>Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA, USA

<sup>3</sup>Duke Clinical Research Institute and Department of Medicine, Duke University School of Medicine, Durham, NC, USA

<sup>4</sup>Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
 <sup>5</sup>Biostatistics Unit, Group Health Research Institute, Seattle, WA, USA

<sup>6</sup>Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA

<sup>7</sup> Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

<sup>8</sup>Division of General Internal Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

#### ABSTRACT

The Mini-Sentinel is a pilot program that is developing methods, tools, resources, policies, and procedures to facilitate the use of routinely collected electronic healthcare data to perform active surveillance of the safety of marketed medical products, including drugs, biologics, and medical devices. The U.S. Food and Drug Administration (FDA) initiated the program in 2009 as part of its Sentinel Initiative, in response to a Congressional mandate in the FDA Amendments Act of 2007.

After two years, Mini-Sentinel includes 31 academic and private organizations. It has developed policies, procedures, and technical specifications for developing and operating a secure distributed data system comprised of separate data sets that conform to a common data model covering enrollment, demographics, encounters, diagnoses, procedures, and ambulatory dispensing of prescription drugs. The distributed data sets currently include administrative and claims data from 2000 to 2011 for over 300 million person-years, 2.4 billion encounters, 38 million inpatient hospitalizations, and 2.9 billion dispensings. Selected laboratory results and vital signs data recorded after 2005 are also available. There is an active data quality assessment and characterization program, and eligibility for medical care and pharmacy benefits is known. Systematic reviews of the literature have assessed the ability of administrative data to identify health outcomes of interest, and procedures have been developed and tested to obtain, abstract, and adjudicate full-text medical records to validate coded diagnoses. Mini-Sentinel has also created a taxonomy of study designs and analytical approaches for many commonly occurring situations, and it is developing new statistical and epidemiologic methods to address certain gaps in analytic capabilities.

Assessments are performed by distributing computer programs that are executed locally by each data partner. The system is in active use by FDA, with the majority of assessments performed using customizable, reusable queries (programs). Prospective and retrospective assessments that use customized protocols are conducted as well. To date, several hundred unique programs have been distributed and executed.

Current activities include active surveillance of several drugs and vaccines, expansion of the population, enhancement of the common data model to include additional types of data from electronic health records and registries, development of new methodologic capabilities, and assessment of methods to identify and validate additional health outcomes of interest. Copyright © 2012 John Wiley & Sons, Ltd.

KEY WORDS-Mini-Sentinel; FDA; U.S. Food and Drug Administration; Sentinel Initiative; FDA Amendments Act of 2007

#### INTRODUCTION

Mini-Sentinel is a collaboration between the U.S. Food and Drug Administration (FDA), 31 academic

Copyright © 2012 John Wiley & Sons, Ltd.

and private organizations, and hundreds of scientists to develop the capability to use routinely collected electronic healthcare data to perform active surveillance of the safety of marketed medical products, including drugs, biologics, and medical devices. FDA initiated the program in 2009 as part of its Sentinel Initiative. The Initiative is a response to a Congressional

<sup>\*</sup>Correspondence to: R. Platt, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA. E-mail: Richard\_Platt@harvard.edu

mandate in the FDA Amendments Act of 2007 to perform active surveillance of the safety of approved drugs through use of routinely collected electronic health information resulting from the care of at least 100 million people.<sup>1,2</sup> The Mini-Sentinel is a pilot program charged with developing the framework, data resources, analytic capabilities, policies, and procedures to satisfy this mandate. In this article, we provide an overview of the Mini-Sentinel's status and direction. Additional information is available in a series of articles describing specific activities<sup>3</sup> and on the Mini-Sentinel website, www.mini-sentinel.org.

#### VISION AND MISSION

The FDA's vision is creation of a system that can use routinely collected electronic health information to support active surveillance of approved medical products, including drugs, biologics, and medical devices, in near real time.<sup>4</sup> Such a system will augment, but not replace, other means of surveillance, including examination of spontaneously reported adverse events. Achieving this vision requires development of a methodologic framework to guide safety surveillance assessments, and creation of the ability to rapidly define cohorts of individuals exposed to medical products of interest, to capture specific health outcomes, and to perform a core set of assessments using customizable computer programs. FDA is committed to achieving this vision through the use of distributed data methods, that is, without creating a centralized data repository.

Mini-Sentinel's mission is to create a "laboratory" that develops and evaluates policies and procedures, organizational structures, and scientific methods that might later be used in a fully operational Sentinel System.<sup>5</sup> Mini-Sentinel activities will thus offer the FDA the opportunity to evaluate safety issues in existing automated health care data systems and to learn more about barriers and challenges to real-time active surveillance using electronic healthcare data.

The initial focus of Mini-Sentinel is on signal refinement, which is the assessment of predefined exposureoutcome pairs to determine whether there is evidence of association. As shown in Figure 1, signal refinement is the second of three steps that begin with signal generation. The exposure-outcome pairs assessed during signal refinement may be identified through signal generation activities using automated data, from the product's clinical development program, through prior knowledge about the product in question or similar products, via spontaneously reported adverse events, or from other sources. Mini-Sentinel is also working on signal generation methods, although this is not a major focus at present.

Mini-Sentinel's signal refinement activities will ordinarily comprise either rapid one-time assessment of the accumulated experience of a product, or prospective repeated (sequential) monitoring of data as it accumulates. In either case, the emphasis of signal refinement is on speed and the use, as much as possible, of standardized methodologic approaches and tools. Signal evaluation, the third step in active surveillance, continues the work of signal refinement, focusing on assessing whether an association is likely to be causal, and addressing questions such as dose-response, duration-response, and inter-individual variability in risk. There is some overlap between the activities of signal refinement and signal evaluation, with the latter typically depending more heavily on customized, in-depth, study-specific protocols. Signal evaluation is not currently a focus of the Mini-Sentinel's activities.

Another Mini-Sentinel activity is rapid assessment of the impact of FDA's regulatory activities. The goal of such assessment is to evaluate the impact of new regulation, such as a new boxed warning, on both prescribing and health outcomes.

Mini-Sentinel's current activities thus include these domains: (i) developing a consortium of data partners and other content experts, (ii) developing policies and procedures, (iii) creating a distributed data system with access to electronic healthcare data and full-text medical records, (iv) developing secure communications capabilities, (v) evaluating extant methods in safety science and developing new epidemiological and statistical methods as needed, (vi) evaluating FDA-identified medical product-adverse event pairs of concern, and (vii) assessing the impact of selected FDA regulatory actions.

#### GOVERNANCE

Mini-Sentinel has developed policies to govern its work.<sup>6</sup> A foundational policy classifies the work of the Mini-Sentinel as public health practice, not research, from the perspective of both the Common Rule that governs research involving human subjects and the Health Insurance Portability and Accountability Act (HIPAA). This classification is the result of determinations by the Department of Health and Human Services' Office for Human Research Protections, with regard to interpretation of the Common Rule, and by FDA, with regard to HIPAA. As a matter of policy, Mini-Sentinel minimizes the transfer of protected health information and proprietary data. The use of a distributed data system plays a central role in implementation

l	Signal Generation Signal Rennement - Signal Evaluation					
Aim = Identify excess risk	All (suspected and unanticipated) adverse events (AEs), all products	Specific product:AE pairs of prior concern	A highly suspected product:AE pair			
Approach	Consider many AEs or Product:AE pairs (100's, 1000's)	Prospective repeated (sequential) monitoring of accumulating data, or one-time expedited analysis of product:AE pairs (typically 5-10)	One-time, in-depth and rigorous investigation of a single pair			

Signal Generation Signal Refinement Signal Evaluation

Figure 1. Stages of postmarket active medical product safety surveillance. The Mini-Sentinel's principal focus is on signal refinement

of this policy. An independent panel of experts in patient privacy assessed the Mini-Sentinel's policies regarding the use of healthcare information.<sup>7</sup>

Additional policies govern the data partners' participation.<sup>8</sup> Key provisions include their status as full partners in the development and implementation of scientific protocols and in interpretation of results, their ability to choose whether or not they participate in specific activities, and their right to use for other purposes their own data that they have transformed into the Mini-Sentinel's common data model format. Mini-Sentinel policies also commit FDA and the investigators to making publicly available the program's policies, tools, methods, protocols, computer programs, and scientific findings. They also address the handling of non-public and confidential information, and conflict of interest.

## THE MINI-SENTINEL DISTRIBUTED DATA SYSTEM

The Mini-Sentinel's principal data resource is a distributed data system comprised of information held by each data partner. Each data partner retains physical and operational control over its own data. This organizational structure has several advantages. It satisfies FDA's requirement that the Mini-Sentinel not establish a centralized data repository, which might raise public concern about potential misuse of confidential medical data. The distributed design avoids the need to create, secure, maintain, and manage access to a complex central data warehouse. It also avoids data partners' concerns about sharing both individuals' confidential information and their own proprietary data. Additionally, it ensures that local content experts maintain a close relationship with the data. This relationship is important because data partners have the best understanding of their data and its uses; valid use and interpretation of findings requires input from the data partners. This knowledge has been critical to understanding appropriate use and interpretation of data, even after its transformation into a common format. Differences in the delivery of care and in coding practices between health plans, and within health plans over time, are typically undocumented and difficult to infer based on data inspection alone. This information is typically only available to individuals with detailed knowledge of a health plan's or practice's operations.

The distributed data system requires each data partner to transform its data to a common data model based on a standard format according to pre-specified definitions. This transformation in advance of use confers two major operational advantages. It allows extensive quality assurance evaluation to assess completeness of the data and identification and remediation of many data quality problems before the data are used to address medical product safety questions. The common data model also allows assessments to be performed through the use of computer programs that are distributed and then executed without sitespecific modification. The use of distributed programs makes highly efficient use of programmer effort and eliminates the potential for protocols to be implemented differently in different systems.

The common data model is comprised of separate tables, each of which contains a specific type of data. This structure is intended to allow the model to evolve to accommodate FDA's needs and the availability of additional data types.<sup>9</sup> The model currently focuses on administrative and claims data. The data areas it encompasses include enrollment, demographics, outpatient pharmacy dispensing, utilization (encounters,

diagnoses, procedures), and mortality (death and cause of death). The model also incorporates clinical data including vital signs, smoking status, and results of ten priority laboratory tests recorded since 2005.

As of July 2011, the distributed dataset contained quality-checked data held by 17 partner organizations. The data covered nearly 100 million individuals (individuals who belonged to more than one participating health plan during the past several years are counted in each plan) for whom there is well-defined eligible person-time during which medically attended events are known. There were over 300 million person-years of observation time, 2.4 billion unique encounters including 38 million acute inpatient stays, and 2.9 billion dispensing of prescriptions. The dataset is refreshed periodically. The development, content, and use of the distributed dataset are described in more detail by Curtis et al.<sup>10</sup> Special considerations for assessment of the safety of vaccines, such as linkage to state immunization registries, are described separately.<sup>11</sup> The Mini-Sentinel's vaccine-related activities are collectively named the Post-licensure Rapid Immunization Safety Monitoring (PRISM) system. PRISM was initiated as a separate single purpose program to evaluate the safety of the H1N1 influenza vaccine; it was then incorporated into the Mini-Sentinel to continue surveillance of influenza and other vaccines.

Data queries (programs) are distributed and returned via a secure portal, as shown in Figure 2. Mini-Sentinel uses three types of queries. It uses a menu-driven query generator for simple questions, such as determining the number of exposures to specific products or the number and age/sex distribution of individuals with a diagnosis or procedure of interest.<sup>12</sup> These queries run against pre-compiled summary tables, thus avoiding the computational overhead involved in analyzing the full distributed dataset. The data partners can also be confident that the queries do not request sensitive information as the tables do not contain personally identifiable information.

For more complex types of recurring queries, Mini-Sentinel uses customizable, reusable (modular) programs.<sup>9</sup> These programs execute in data partners' full distributed datasets. An example is a program that



Figure 2. Querying the Mini-Sentinel distributed database. Each query involves five steps: 1) A query (program) is created and then posted by an authorized user on the secure portal. 2) Data partners are notified and retrieve the query. 3) Data partners review the query and execute it against their local data. 4) Data partners review the results, which are typically counts, e.g., number of exposed individuals, amount of exposed person-time, number of individuals with outcomes of interest. 5) Data partners submit their results using the secure portal. 6) The results are reviewed and then combined with other data partners' results

identifies cohorts of new users of specific products, determines the number of dispensings, the amount of exposed person-time, and the number of outcomes of interest observed during exposed time. These reusable programs allow users to specify parameters such as inclusion and exclusion criteria, and the new user and outcome definitions. These programs carry several operational advantages, including the fact that the programs are extensively vetted to assure that they perform the desired assessments, and that they execute efficiently in the data partners' diverse computing environments. The program logic is pre-approved by the data partners so the output generated requires minimal evaluation by data partners. These programs produce counts, and in some cases, rates, for specified age, sex, and calendar time strata, but do not currently adjust for confounding factors. The third type of query involves custom programs that perform assessments beyond the scope of existing modular programs. These are typically used to support prospective surveillance protocols, which may have unique needs. Mini-Sentinel attempts to capture the novel programming performed for these studies and make it available through a program library or by incorporating it into a new modular program.

#### METHODS DEVELOPMENT

Mini-Sentinel investigators have developed a taxonomy of study designs to guide the development of active surveillance protocols and also of new modular programs.<sup>13</sup> This taxonomy considers various combinations of exposure attributes (e.g., acute, chronic), outcome attributes (e.g., rare, common), and characteristics of the exposure-outcome relationship, with the intent of expediting the choice of study design aspects for a wide range of exposures and outcomes. The taxonomy continues to evolve to include considerations of analytic strategy and conditions specific to assessment of adverse reactions to vaccines.

Substantial effort has also been devoted to clarifying the applicability of semi-automated methods for control of confounding in cohort designs, such as the high-dimensional propensity score,<sup>14</sup> and to providing guidance regarding the strengths, limitations, and practicability of case-only methods.<sup>15</sup> Mini-Sentinel investigators also tested a multivariable adjusted selfcontrolled case series and conducted statistical simulation studies on aspects of semi-automated covariate identification and selection strategies.<sup>16,17</sup>

Because a substantial portion of the Mini-Sentinel portfolio will involve prospective repeated (sequential) assessment of accumulating data for specific exposures and outcomes, Mini-Sentinel investigators have begun to explore the challenges associated with applying sequential designs in observational safety surveillance settings.<sup>18</sup> To date, sequential testing methods have primarily been used in randomized clinical trials. Although their application in observational contexts like Mini-Sentinel is promising, several issues that are generally not of concern (or are of much smaller magnitude) in trials complicate matters. These include (i) lack of experimental control, which can vield confounding, unpredictable new user accrual rates and composition over time, missing data, and misclassification, (ii) heterogeneous sites contributing data in a distributed environment that prevents individuallevel data pooling and thus constrains analytic options, and (iii) the safety outcomes typically evaluated can be rare, which introduces instability and may require small sample testing strategies. In addition, the scientific and regulatory aims for postmarket safety, which inherently impact key sequential design decisions such as the frequency of interim testing, are different than in premarket trials and require additional consideration.

Recognizing the need for better ability to choose between different approaches to sequential assessments in observational safety surveillance settings, Cook and colleagues performed simulations to compare the performance of four methods, which each use a different confounder adjustment strategy: the Lan-Demets group sequential error spending approach, a group sequential likelihood ratio test, the conditional sequential sampling procedure, and a group sequential generalized estimating equations approach.<sup>19</sup> The simulation evaluated type 1 error rate, power, and time-to-signal detection, under varying assumptions about outcome prevalence, exposure, and confounder complexity.

#### USING AND INTERPRETING THE DATA

Mini-Sentinel investigators have devoted considerable effort to understanding the state of knowledge in use of administrative data to identify the health outcomes of greatest interest as endpoints for safety assessments of medical products, and the validity of current methods to identify outcomes. In collaboration with FDA, investigators identified the 20 highest priority outcomes among a candidate list of 140 outcomes for which there had been no recent review. Investigators then performed systematic reviews of these 20 conditions, drawing on protocols that had been developed by the Observational Medical Outcomes Partnership.<sup>20</sup> The methods for conducting these reviews have been summarized by Carnahan and Moores, along with lessons learned about the strengths and limitations of the review process.<sup>21</sup> A high-level classification of the findings of these reviews is provided in Table 1. Carnahan and Moores also identified the gaps in our knowledge of the usefulness of administrative data for identifying these outcomes and offered suggestions for additional research in this area.<sup>22</sup>

Our expectation is that instances of potential outcomes identified through use of administrative data will usually require review and adjudication of fulltext medical records in situations that require the predictive value of designation as a case to be very high. Confirmation might be needed if signal refinement discovers evidence of excess risk. Cutrona and colleagues describe the process Mini-Sentinel developed to identify cases of acute myocardial infarction using distributed programs, to have data partners obtain the

Table 1. Utility of administrative data to identify health outcomes of interest

Good utility*	Moderate utility <sup>†</sup>	Little utility <sup>‡</sup>		
Cerebrovascular accident and transient ischemic attacks <sup>25</sup>	Atrial fibrillation <sup>26</sup>	Anaphylaxis <sup>27</sup>		
Heart failure <sup>28</sup>	Ventricular arrhythmia <sup>29</sup>	Hypersensitivity reactions other than anaphylaxis <sup>30</sup>		
Venous thromboembolism <sup>31</sup>	Seizures, convulsions, or epilepsy <sup>32</sup>	Seizures, convulsions, or epilepsy <sup>32</sup>	Erythema multiforme and other serious skin reactions <sup>33</sup>	
Angioedema <sup>27</sup>	Depression <sup>34</sup>	Acute respiratory failure <sup>35</sup>		
Revision of total hip arthroplasty <sup>36</sup>	Pancreatitis <sup>37</sup>	Pulmonary fibrosis and interstitial lung disease <sup>38</sup>		
		Infection related to blood products, tissue grafts, or organ transplantation <sup>39</sup>		
		Transfusion-associated sepsis <sup>40</sup>		
		Transfusion reaction caused by ABO incompatibility <sup>41</sup>		
		Suicide – attempted or completed $^{42}$		
		Revision of knee arthroplasty <sup>36</sup>		
		Lymphoma <sup>43</sup>		

\*Positive Predictive Values consistently >70% to identify acute or incident events across most of multiple studies examining relatively generalizable study populations.

Positive Predictive Values 50–70%, inconsistent findings, based on few studies, limited information on identifying acute or incident events, sensitivity of algorithms questionable, or limited generalizability based on study populations.

<sup>\*</sup>Positive Predictive Values <50%, very limited or dated information on validity of algorithms compared to medical record review, or other substantial limitations in algorithm performance or evidence.

Copyright © 2012 John Wiley & Sons, Ltd.

relevant portions of full-text inpatient medical records, and provide either redacted records or abstracted information to an expert panel for adjudication.<sup>23</sup> Notably, it was possible to obtain redacted information from 93% (143/153) of requested full-text records.

# ASSESSMENT OF THE SAFETY OF MEDICAL PRODUCTS

The Mini-Sentinel distributed dataset became usable for distributed queries in early 2011. To date, the data partners have executed several hundred distributed programs in response to FDA queries. Examples of modular program queries included assessment of the occurrence of acute myocardial infarction or stroke among new users of drugs used to treat Parkinson's disease, celiac disease among recipients of angiotensin receptor blockers, and cardiac outcomes among individuals who were dispensed prescription drugs for smoking cessation.

One-time protocol-based assessments include initiation of assessments of intussusception after two rotavirus vaccines, and venous thromboembolism following human papilloma virus vaccine.<sup>11</sup> A prospective sequential evaluation of the occurrence of acute myocardial infarction among users of different antidiabetic drugs is also in progress.<sup>24</sup>

#### NEXT STEPS

Near-term objectives include expanding the number and type of assessments, increasing the size and diversity of the covered population, including data from ambulatory and inpatient electronic health records and registries, and broadening the range of medical products and outcomes under observation. Additional data from two large national health plans are expected to become available within the next year, substantially increasing the size of the population. Expansion of available laboratory results and development of modular programs that incorporate height, weight, blood pressure, smoking status, and outpatient laboratory test results in conjunction with drug exposures and clinical diagnoses are planned. Algorithms will be developed to identify populations of special interest, such as pregnant women and patients with renal dysfunction. The availability of information about exposures to blood products will be explored.

Ongoing and planned methodologic studies include evaluation of inverse probability weighting to adjust for confounding within a sequential monitoring framework, evaluation of methods for anonymous linkage of individuals who are represented in more than one data source, methods for distributed multivariable-adjusted analysis, assessing the roles of propensity score and disease risk score methods in monitoring the safety of new medical products, additional simulation capabilities, and work on signal generation.

Systematic reviews of the validity of coded diagnoses for additional health outcomes of interest that are especially relevant to evaluation of vaccine safety will be performed. Validation studies that involve adjudication of full-text medical records will be performed for severe acute liver injury, venous thromboembolism, intussusception, and anaphylaxis.

Surveillance activities will include new prospective and retrospective assessments with customized protocols, as well as assessment of the impact of regulatory action.

#### LONGER TERM

Developing a robust system for active surveillance of medical product safety is a long-term, complex initiative. It will be necessary to implement it in stages as scientific methods and data infrastructure mature. Ongoing effort will be required to achieve an appropriate balance between the need for timeliness in assessing the safety of medical products and avoiding misleading conclusions. It will also be necessary to ensure privacy and security within the distributed system and to continue to address the concerns of stakeholders including patients and the public. Finally, it will be important to consider ways in which the resources and methods that the Mini-Sentinel develops can serve as a national resource to support other secondary uses of electronic health data, including clinical effectiveness and quality of care.

#### CONFLICT OF INTEREST

The authors have declared that there is no conflict of interest.

#### **KEY POINTS**

- Mini-Sentinel has created a distributed data network, analytic methods, and policies to enable use of routinely collected electronic health information to assess the safety of marketed medical products
- The network is currently in routine use by FDA
- Mini-Sentinel focuses on rapid assessment of past experience, prospective assessment of accumulating data, and assessment of changes in utilization and health outcomes after regulatory action
- This network has the potential to address national needs beyond safety of medical products.

#### ACKNOWLEDGEMENTS

Mini-Sentinel is funded by the Food and Drug Administration through Department of Health and Human Services Contract Number HHSF223200910006I.

#### REFERENCES

- Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System--a national resource for evidence development. *N Engl J Med* 2011; 364: 498–499.
- Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network--improving the evidence of medical-product safety. N Engl J Med 2009; 361: 645–647.
- Platt R, Carnahan R (eds). The U.S. Food and Drug Administration's Mini-Sentinel Program: status and direction. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 1–8.
- Robb MA, Racoosin JA, Sherman RE *et al.* FDA's Sentinel Initiative: Expanding the horizons of medical product safety. *Pharmacoepidemiol Drug Saf* 2012; **21**(S1): 9–11.
- The Sentinel Initiative: Access to electronic healthcare data for more than 25 million lives: Achieving FDAAA Section 905 Goal One. U.S. Food and Drug Administration, 2010. http://www.fda.gov/downloads/Safety/FDAsSentinelInitiative/ UCM233360.pdf [18 October 2011].
- Forrow S, Campion DM, Herrinton LJ et al. The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 12–17.
- McGraw D, Rosati K, Evans B. A policy framework for public health uses of electronic health data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 18–22.
- Mini-Sentinel Principles and Policies, 2011. www.minisentinel.org/work\_ products/About\_Us/Mini-Sentinel-Principles-and-Policies-May-2011.pdf [19 October 2011].
- Data Activities. Mini-Sentinel. http://mini-sentinel.org/data\_activities/ [22 October 2011].
- Curtis LH, Weiner MG, Boudreau DM et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol Drug* Saf 2012; 21(S1): 23–31.
- Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 291–297.
- Mini-Sentinel Distributed Query Tool: System Description and Technical Documentation. www.mini-sentinel.org/data\_activities/details.aspx?ID=116 [24 October 2011].
- Gagne JJ, Fireman B, Ryan PB et al. Design considerations in an active medical product safety monitoring system. Pharmacoepidemiol Drug Saf 2012; 21(S1): 32–40.
- Rassen JA, Schneeweiss S. Using high dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 41–49.
- Maclure M, Fireman B, Nelson JC *et al.* When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 50–61.
- Mini-Sentinel Methods Development: Case-Based Methods Workgroup Report, 2011. www.mini-sentinel.org/work\_products/Statistical\_Methods/Mini-Sentinel\_ Methods\_Case-Based-Report.pdf [17 October 2011].
- Myers J, Rassen J, Gagne J et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. Am J Epidemiol 2011; 174: 1223–1227.
- Nelson J, Cook A, Yu O et al. Challenges in the design and analysis of sequentially-monitored postmarket safety surveillance evaluations using electronic observational health care data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 62–71.
- Cook AJ, Tiwari RC, Wellman RD *et al.* Statistical approaches to group sequential monitoring of postmarket safety surveillance data: Current state of the art for use in the Mini-Sentinel pilot. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 72–81.
- Health Outcomes of Interest Library. Observational Medical Outcomes Partnership. http://omop.fnih.org/HOI [17 October 2011].
- Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 82–89.
- Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 90–99.
- Cutrona SL, Toh S, Iyer A et al. Design for validation of acute myocardial infarction cases in Mini-Sentinel. Pharmacoepidemiol Drug Saf 2012; 21(S1): 274–281.
- Fireman B, Toh S, Butler MG et al. A protocol for active surveillance of acute myocardial infarction in association with the use of a new antidiabetic pharmaceutical agent. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 282–290.
- 25. Andrade SE, Harrold LR, Tjia J et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic

Pharmacoepidemiology and Drug Safety, 2012; 21(S1): 1–8 DOI: 10.1002/pds attack using administrative data. *Pharmacoepidemiol Drug Saf* 2012; **21**(S1): 100–128.

- Jensen PN, Johnson K, Floyd J, Heckbert SR, Carnahan R, Dublin S. A systematic review of validated methods for identifying atrial fibrillation using administrative data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 141–147.
- Schneider G, Kachroo S, Jones N et al. A systematic review of validated methods for identifying anaphylaxis, including anaphylactic shock and angioneurotic edema, using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 240–247.
- Saczynski JS, Andrade SE, Harrold LR et al. A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 129–140.
- Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying ventricular arrhythmias using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 148–153.
- Schneider G, Kachroo S, Jones N et al. A systematic review of validated methods for identifying hypersensitivity reactions other than anaphylaxis (fever, rash, and lymphadenopathy), using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 248–255.
- Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying venous thromboembolism using claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 154–162.
- Kee VR, Gilchrist B, Granner MA, Sarrazin NR, Carnahan RM. A systematic review of validated methods for identifying seizures, convulsions, or epilepsy using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 183–193.
- 33. Schneider G, Kachroo S, Jones N *et al*. A systematic review of validated methods for identifying erythema multiforme major/minor/not otherwise specified, Stevens-Johnson Syndrome, or toxic epidermal necrolysis using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; **21**(S1): 236–239.
- Townsend L, Walkup JT, Crystal S, Olfson M. A systematic review of validated algorithms for identifying depression using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 163–173.

- Jones N, Schneider G, Kachroo S, Rotella P, Avetisyan R, Reynolds MW. A systematic review of validated methods for identifying acute respiratory failure using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 261–264.
- Singh JA, Kundukulam JA, Bhandari M. A systematic review of validated methods for identifying orthopedic implant removal and revision using administrative data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 265–273.
- Moores K, Gilchrist B, Carnahan R, Abrams T. A systematic review of validated methods for identifying pancreatitis using administrative data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 194–202.
- Jones N, Schneider G, Kachroo S, Rotella P, Avetisyan R, Reynolds MW. A systematic review of validated methods for identifying pulmonary fibrosis and interstitial lung disease using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 256–260.
- Carnahan RM, Moores KG, Perencevich EN. A systematic review of validated methods for identifying infection related to blood products, tissue grafts, or organ transplants using administrative data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 213–221.
- Carnahan RM, Herman RA, Moores KG. A systematic review of validated methods for identifying transfusion-related sepsis using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 222–229.
- Carnahan RM, Kee VR. A systematic review of validated methods for identifying transfusion-related ABO incompatibility reactions using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 230–235.
- Walkup JT, Townsend L, Crystal S, Olfson M. A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 174–182.
- Herman RA, Gilchrist B, Link BK, Carnahan R. A systematic review of validated methods for identifying lymphoma using administrative data. *Pharmacoepidemiol Drug Saf* 2012; 21(S1): 203–212.

8

# RESEARCH

# BMJ

## Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings

Richard L Tannen, professor of medicine, Mark G Weiner, associate professor of medicine, Dawei Xie, assistant professor of biostatistics and epidemiology

#### ABSTRACT

of Medicine, 295 John Morgan Building, 36th and Hamilton Walk, Philadelphia, PA 19104, USA Correspondence to: R L Tannen tannen@mail.med.upenn.edu

University of Pennsylvania School

Cite this as: *BMJ* 2009;338:b81 doi:10.1136/bmj.b81 **Objectives** To determine whether observational studies that use an electronic medical record database can provide valid results of therapeutic effectiveness and to develop new methods to enhance validity.

**Design** Data from the UK general practice research database (GPRD) were used to replicate previously performed randomised controlled trials, to the extent that was feasible aside from randomisation.

**Studies** Six published randomised controlled trials. **Main outcome measure** Cardiovascular outcomes analysed by hazard ratios calculated with standard biostatistical methods and a new analytical technique, prior event rate ratio (PERR) adjustment.

**Results** In nine of 17 outcome comparisons, there were no significant differences between results of randomised controlled trials and database studies analysed using standard biostatistical methods or PERR analysis. In eight comparisons, Cox adjusted hazard ratios in the database differed significantly from the results of the randomised controlled trials, suggesting unmeasured confounding. In seven of these eight, PERR adjusted hazard ratios, whereas in five they didn't differ significantly, and in three were more similar to the hazard ratio from the randomised controlled trial, yielding PERR results more similar to the randomised controlled trial than Cox (P<0.05).

**Conclusions** Although observational studies using databases are subject to unmeasured confounding, our new analytical technique (PERR), applied here to cardiovascular outcomes, worked well to identify and reduce the effects of such confounding. These results suggest that electronic medical record databases can be useful to investigate therapeutic effectiveness.

#### **INTRODUCTION**

The future widespread implementation of electronic records in clinical practice will provide an enormous opportunity for research related to medical treatments, provided this information is compiled into robust, well designed databases and analysed with appropriate methods. By contrast, incorrect analyses could have important negative effects on medical treatment and health policy. Therefore, before implementation of this approach for assessing effectiveness of treatment, we need to assess the validity of the results from studies using such databases and of the study design and analytical strategies that are most likely to yield valid results. The need for further investigation into these strategies is widely supported.<sup>1-7</sup>

Two major potential problems could arise in the use of medical record databases to provide reliable information concerning treatment outcomes: the quality of the data contained within the database and the ability of analyses of observational—that is, nonexperimental—data to provide valid results.

Considerable controversy exists over whether observational studies can provide reliable information on effectiveness of therapeutics.<sup>128-15</sup> Because of their ability to balance measured and unmeasured confounders, randomised controlled trials remain the highest level of evidence, whereas the quality of evidence from observational studies is lower because of confounding by indication and other biases related to the effects of unmeasured covariates. Several comparative analyses suggest that observational studies often yield results reasonably consistent with those of randomised controlled trials. Nevertheless, there are several well documented examples where the results from observational studies were misleading.1-3781617 Hormone treatment to protect against coronary artery disease in postmenopausal women is one highly cited example.<sup>7-20</sup> Some authorities believe that the results of observational studies should be an important component of evidence based medicine; some suggest their reliability is limited to conditions where confounding by indication is unlikely, as, for example, in studies of unanticipated adverse effects of drugs,<sup>21</sup> whereas others are sceptical of their value.  $^{^{1\,2\,7\text{-}15}}$ 

An important limitation applicable to previous comparative analyses is that most of the observational studies did not have rigorous inclusion and exclusion



Fig 1 | Selection process for participants in database studies

criteria, exposure definitions, and outcomes identical to the randomised controlled trials so that lack of randomisation was not the only important difference.<sup>12152223</sup>

To overcome these limitations in validating an observational study, we tested the value of a comprehensive longitudinal electronic clinical database, the UK General Practice Research Database (GPRD), using studies designed to replicate the design of previously performed randomised controlled trials to the extent that was feasible aside from randomisation.<sup>24</sup> Validity of the method was measured by comparing the outcomes of the replicated GPRD study with those of the randomised controlled trial.<sup>25-29</sup> The GPRD study results depended on both the quality of the information in the database and whether observational data can reproduce results from a randomised controlled trial.

We examined both the potential research value of the electronic medical record database and the validity of observational studies. We also used a new analytical method, prior event rate ratio (PERR) adjustment, to enhance the validity of the results.

#### **METHODS**

#### GPRD database

The UK GPRD database contains information from the electronic medical records of primary care practices encompassing a representative sample of about 5.7% of the UK population during 1990-2000 and contains records of over eight million patients.<sup>2425</sup> It includes the complete primary care medical record, comprehensive information on essentially all medications prescribed, and information from outside consultants and admissions to hospital. The box details limitations and advantages of the database.

#### GPRD study protocol

Table 1 summarises database replications of six randomised controlled trials that have been performed and reported in detail elsewhere.<sup>20 30-34</sup> As far as possible the database studies used the same inclusion and exclusion criteria, a similar study time frame, and a similar treatment regimen as the randomised trials.<sup>25-29</sup> Thus the major primary difference was the lack of randomisation in the database studies, albeit that other

issues such as use of placebo, nature of healthcare delivery, and some characteristics of subjects entered into randomised trials compared with those existing in the general population can differ between a randomised trial and a database study.

Selection of the subjects for inclusion in the database studies followed the outline shown in figure 1. First the exposed cohort was selected from all database subjects who met the inclusion criteria and received treatment with the study treatment during a predefined recruitment interval. The exposed cohort was finalised after elimination of patients with exclusion criteria. Their start time was the day of the first prescription of the study drug. The unexposed cohort was selected from all patients who met the inclusion criteria but did not receive the study drug during the recruitment interval. They were then age and sex matched to the exposed patients with a computerised random selection program, and their start time was considered identical to that of the matched exposed patients. Then, those who had exclusion criteria were eliminated.

The selection process differed for the database matched to the Syst-Eur study because study entry and start time for both the exposed and unexposed cohorts was determined by measured blood pressure that indicated systolic hypertension.<sup>25</sup>

All database studies ended on a predefined date or on outcome stop points defined in the randomised controlled trial. Patients were considered lost to follow-up if they left the practice or the practice was eliminated from the database before the end date. We analysed database studies using a simulated "intention to treat" paradigm where subsequent treatment of the exposed and unexposed patients did not modify study end time and also an "as treated" analysis in which the study ended for an exposed or unexposed patient who deviated from their treatment protocol.

#### Statistical analysis

We determined Cox unadjusted and adjusted hazard ratios for all outcomes. The adjusted hazard ratios used a predetermined set of potential confounders that included key demographics, medications at baseline, and identified medical conditions. We imputed missing values for systolic blood pressure, body mass index, and smoking<sup>35</sup> and created five separate datasets. The final estimates combined the results from the five datasets, as described previously.<sup>2635-37</sup>

We also analysed results with a propensity score approach, which used all demographics, drug use at baseline, and identified medical conditions as confounders.<sup>25 26 38-40</sup> Propensity scores were estimated using logistic regression with the outcome being the indicator of treatment and the covariates being all confounders considered. For those with no missing data, all covariates were used; whereas for those with missing data (for body mass index, systolic blood pressure, or smoking), we used separate logistic regression models, which excluded the missing covariates, to estimate propensity scores. Analysis stratified by the propensity scores balances the treated and untreated groups with respect to the observed covariates used in estimating the propensity scores. We determined outcome hazard ratios for each fifth of the propensity score and combined the five hazard ratios to determine an overall hazard ratio using a Cox model treating the fifths as strata with different baseline hazards. The propensity score thereby accounts for missing confounders in a different fashion from the multiple imputation method used with the Cox analysis. The matched database study for Syst-Eur, our first study, was analysed only with propensity score analysis.<sup>25</sup>

We also used a prior event rate ratio (PERR) approach to adjust the Cox hazard ratio, as described recently.<sup>2829</sup> This analysis requires that neither the exposed nor unexposed patients are treated with the study drug before the start of the study. It assumes that the hazard ratio of the exposed to unexposed for a specific outcome before the start of the study reflects the combined effect of all confounders (both measured and unmeasured) independent of any influence of treatment.

To apply the PERR adjustment method, we divided the unadjusted hazard ratio of exposed versus unexposed groups during the study by the unadjusted hazard ratio of exposed versus unexposed "before" the study. Thus if p=prior events and s=study events, the calculation is: PERR adjusted HR=HRs/HRp. We obtained confidence intervals for the PERR adjusted hazard ratio using a bootstrap technique.<sup>28</sup> Hazard ratios are reported because of variable observation times for patients both before and during the study; though incidence rate ratios produced similar results.

In all studies we carried out the PERR analysis using a subset of patients who did not take the study drug at any time before the start of the study. In no instance did Cox adjusted hazard ratios for this subset differ meaningfully from the results in the overall cohort.

The time interval used to assess previous events encompassed 1 January 1987 to the patient's start time. If a patient had no medical or treatment record before that date, their time interval began on the earliest subsequent date with a record. If they had no records before the study start time, they were not used in this analysis. The average time of the previous period for all the outcomes assessed averaged 3.52 years (range 2.8-3.9 years). Analysis of the impact of the duration of the previous time period using the empirical data in these studies suggested that encompassing events from 3-4. 5 years before study start time did not meaningfully influence the results of the PERR analysis.

We compared differences between the hazard ratio from the randomised trial and the database using a standard normal z test, where the z score was obtained from the difference between the logarithm of the hazard ratio divided by the standard error of that difference.<sup>25</sup>

#### RESULTS

We collated and analysed the collective results of six database studies reported previously.<sup>25-29</sup>

## Comparability between replicated database study and randomised controlled trial

The size of the unexposed group in the database study was always larger than the placebo group of the randomised controlled trials (table 1). The exposed group in the database study, however, was smaller than the treated cohort in half the randomised controlled trials. Furthermore, the database was inadequate to replicate several randomised controlled trials because of an insufficient number of exposed patients.

Table 1 Comparison of study characteristics in randomised controlled trials and general practice research database (GPRD)

	No of subjects				Treatment protocol		
	Trials		Dat	tabase			
Study	Treated Placebo		Exposed Unexposed		Trials	Database	
Syst-Eur*	2398	2297	2815	13 956	Nitrendipine, enalapril or hydrochlorothiazide. Target: 20 mm Hg decrease SBP	Dihydropyridine calcium channel blocker, ACE inhibitor, or thiazide diuretic. No target	
WHI (Women's Healt	h Initiative):						
Intact uterus†	8506	8902	13 658	37 730	Conjugated oestrogen 0.625 mg/day; medroxyprogesterone 5.0 mg/day	Conjugated oestrogen 0.625 mg/day; norgestrel 150 μg on days 17-28	
Hysterectomy‡	5310	5429	6890	11 572	Conjugated oestrogen 0.625 mg/day	Conjugated oestrogen 0.625 mg/day	
4S§	2221	2223	1280	2871	Simvastatin 20 mg/day. Target: total cholesterol 115-200 mg/dl	Any statin (80% received simvastatin). No target	
HOPE¶	4645	4652	2812	26 286	Ramipril 10 mg/day	Any ACE I inhibitor with ramipril equivalent >4 mg/day (average dose 6.8 mg/day)	
EUROPA**	6110	6108	2668	12 705	Perindopril 8 mg/day (ramipril equivalent 10 mg/day)	Any ACE inhibitor with ramipril equivalent >4 mg/day (average dose 6.5 mg/day)	

ACE=angiotensin converting enzyme.

\*Antihypertensive treatment of older patients with isolated systolic hypertension in Europe.<sup>25 30</sup>

†Postmenopausal women treated with combined hormone replacement.<sup>20 26 41</sup>

 $\pm$ Postmenopausal women with previous hysterectomy treated solely with conjugated oestrogen <sup>27 31 41</sup>

Scandinavian simvastatin survival study of hypercholesterolaemic subjects with coronary artery disease treated with simvastatin.<sup>28 32</sup>

[Heart outcomes prevention evaluation study of ramipril (angiotensin converting enzyme inhibitor), treatment of patients either with established or at high risk for coronary artery disease.<sup>29</sup>

\*\*European trial on reduction of cardiac events with perindopril (angiotensin converting enzyme inhibitor) in patients with stable coronary artery disease.<sup>29 34</sup>



Fig 2 | Comparisons between hazard ratios from randomised controlled trials (RCT) and adjusted hazard ratios for respective database studies. Data plotted as natural logarithms, so 0 on x axis indicates no difference between exposed and unexposed cohort. Database adjusted hazard ratios shown with both Cox and prior event rate ratio (PERR) adjustment analysis. Results are shown for myocardial infarction, stroke, and coronary revascularisation (CABG/PTCA). GPRD=general practice research database

Although entry criteria were similar for the database studies and randomised trials, the database cohort typically differed from the respective trials in their baseline demographic characteristics, existing comorbidities, and use of cardiovascular drugs.<sup>25-29</sup>

The database treatment protocol precisely replicated the trial in only one study (WHI-hysterectomy (see table 1). The other database studies used the same class of drug, rather than specific drug used in the trial. Furthermore, identical dosing regimens could not be replicated. It is worth noting that the prescription database in GPRD can actually track data on medication prescribing better than many randomised controlled trials.

Finally in contrast with the randomised controlled trials, where randomisation resulted in similar baseline health profiles of the treated and placebo arms, all the database studies except Syst-Eur exhibited differences in the baseline characteristics of the exposed and unexposed groups.

## Comparison of outcomes in the database studies and randomised controlled trials

We focused on randomised controlled trials with primary cardiovascular outcomes because they could be replicated reasonably without the need for laboratory data. We report on death, myocardial infarction, stroke, and coronary revascularisation (coronary artery bypass grafts or percutaneous transluminal coronary angioplasty (CABG/PTCA) or both). These cardiovascular outcomes should be the least susceptible to misclassification errors. Other outcome results are provided in the primary publications, and the results for breast cancer, colon cancer, and hip fracture were similar in both the "intact uterus" and "hysterectomy" WHI randomised controlled trials and their respective database studies.<sup>25-29</sup>

Table 2 and figure 2 show cardiovascular outcomes and statistical comparisons for the six database studies and trials. We have shown simulated "intention to treat" results, but results of the "as treated" analyses did not differ meaningfully. Cox adjusted and prior event rate ratio (PERR) adjusted hazard ratios (performed in five studies) are also shown.

Propensity score analyses (table 3) did not differ meaningfully from the analysis with Cox adjusted hazard ratios. A minor exception was the death outcomes in the HOPE and EUROPA studies, where the propensity score adjusted hazard ratios were slightly lower than the Cox adjusted hazard ratios and slightly more similar to the hazard ratios from the randomised controlled trial.

Results from the WHI randomised controlled trial for the entire cohort and also subdivided by age were reported.<sup>203141</sup> We compared the database studies to the overall WHI randomised controlled trial and also to the results restricted to women aged <70, an age profile more comparable with the study cohorts in the database.

#### Cardiovascular outcomes

In nine of 17 comparisons of cardiovascular outcomes there was no significant difference between the Cox
adjusted hazard ratios from the database and the hazard ratios from the randomised controlled trials (see table 2, which compares the trial hazard ratio, the database Cox adjusted hazard ratio, and the database-PERR adjusted hazard ratios). In none of these nine comparisons did the PERR analysis differ significantly from either the trial hazard ratios or the Cox adjusted hazard ratios.

In eight of the 17 comparisons, however, the Cox adjusted hazard ratios differed significantly from the trial hazard ratios, suggesting the presence of unmeasured confounding. In seven of these eight instances the PERR adjusted hazard ratios differed significantly from the Cox adjusted hazard ratios, and either did not differ significantly (five outcomes) or were more similar (two outcomes) to the trial hazard ratio. In the other outcome the PERR hazard ratio was more similar to the trial but did not differ significantly from the Cox adjusted hazard ratio. A Wilcoxon signed rank test showed that when the Cox adjusted hazard ratio differed significantly from the trial hazard ratio (n=8), the PERR adjusted hazard ratio was significantly (P < 0.05) more similar to the trial hazard ratio than the Cox adjusted hazard ratio.

As the 17 outcomes analysed came from six studies, it is reasonable to question the analysis of each outcome as an independent data point. As shown in the 4S study, however, the two outcomes (myocardial infarction and coronary revascularisation) clearly behaved independently of one another. Unmeasured confounding affected revascularisation but had no discernible effect on myocardial infarction. As the PERR analysis is outcome specific and derived entirely in that fashion, it seems reasonable to analyse the data assuming that each individual outcome is independent.

In the aggregate, when the outcome results from the database studies analysed by conventional statistical methods are confirmed or corrected by the PERR method, they are largely comparable with the results from the respective randomised controlled trials.

The large confidence intervals in the PERR analysis of all the WHI outcomes, which limits the interpretation of this data, were due to the small number of previous events. Surprisingly, despite this limitation,

### The GPRD database

#### Advantages

- Comprehensive national healthcare system
- Representative sample of entire population
- · All care centralised in general practice record
- All medications prescribed by general practitioner, generated by computer
- Contains around eight million patients

## Limitations

- Lacks direct link to laboratory data (laboratory data inadequate)
- Missing data on smoking, systolic blood pressure, body mass index (about 30%)
- Limited data on onset of menopause
- Limited data on admission to hospital
- · Lacks direct link to death certificates (cause of death not reliable)

the PERR adjusted hazard ratio was significantly higher than the Cox adjusted hazard ratio and not different from the randomised controlled trial hazard ratio for both the myocardial infarction and stroke outcomes in the WHI-hysterectomy study, suggesting the presence of unmeasured confounding. The likelihood that unmeasured confounding influenced these two outcomes is consistent with the significant difference between the Cox adjusted hazard ratios and the randomised controlled trial hazard ratios.

## Death

We have shown only Cox adjusted hazard ratios for death because PERR adjustment cannot be done. The Cox and the propensity score adjusted hazard ratios for death resembled the randomised controlled trial results in three studies; however, they were higher than the trial in the Syst-Eur study and lower in both the WHI studies. The WHI results on death should be interpreted cautiously because in both studies a subset of the overall cohort that was not missing any data on baseline body mass index, systolic blood pressure, or smoking did not show a significant decrease in death, despite results comparable with the overall cohort for all other outcomes.

## DISCUSSION

Despite its shortcomings, this careful, albeit not exhaustive, comparison between randomised controlled trials and observational studies using data from an electronic primary care medical record database reveals several important insights. From an overall perspective, our results suggest that observational studies using databases might produce valid results concerning the efficacy of cardiovascular drug treatments.

### Rigour of database studies

Our studies comparing performance of the database and randomised controlled trials were performed in as rigorous fashion as possible.

In addition to using similar inclusion and exclusion criteria and relatively similar time frames, we analysed studies with both a simulated "intention to treat" and an "as treated" design. We analysed data with multiple imputation plus Cox adjusted hazard ratios, and also propensity score plus stratified Cox unadjusted hazard ratios. The propensity score is useful to identify heterogeneity and also incorporates missing data into the analysis in a fashion different from the multiple imputations used with the primary Cox method. We used a subset of the overall cohort without "missing data" on the key confounders (systolic blood pressure, body mass index, and smoking) as a secondary verification analysis to ensure that missing data did not influence the results in the overall cohort. We assessed use of non-study drugs to confirm that cointervention during the study did not account for the results. Computerised random matching and thereby start time delineation for the unexposed group obviated the potential for unanticipated bias related to start time in the unexposed group.

## Overall study results

We analysed results of the outcomes for myocardial infarction, stroke, coronary revascularisation, and death for six comparative studies (table 2 and fig 2). We examined the aggregate database study results with conventional biostatistical analyses (Cox adjusted hazard ratios or propensity score analyses, or both) and our newly described prior event rate ratio (PERR) adjustment technique.<sup>28 29</sup>

When analysed with conventional biostatistical analyses, the database outcome results (independent of death) did not differ significantly from those in the randomised controlled trial in nine of the 17 comparisons. In no instance did the PERR analysis differ significantly from the randomised controlled trial, when there was no difference between the conventional analyses and the trial.

As shown in table 2 and figure 2, when the database outcomes analysed with conventional biostatistical techniques differed significantly from the trial, the PERR analysis results were either not significantly different from or much more similar to the trial results. The instances where the database results analysed by conventional biostatistical methods differed importantly from the results in the trial presumably reflect unmeasured confounding by indication in the database studies. Thus our findings support concerns that the validity of observational studies must always be viewed with circumspection. The studies reported herein, however, suggest that the PERR technique can identify (by differing from the results with standard statistical methods) and largely correct for the effects of unmeasured confounding, when it exists. The availability in the database of previous event rates, rather than only prevalence data, permitted performance of this analysis.

## PERR analytical technique

The underlying hypothesis of the PERR analytical technique is that a comparison between the event rate for a specific outcome in a cohort's exposed and unexposed patients before entry into the study should reflect the effect of all confounders on that specific outcome independent of the effect of treatment. This assumption holds only when neither the exposed nor unexposed patients have been treated with the study drug before the start of the study. If so, the ratio between the previous events in the exposed and

#### Table 2 | Comparison of outcome hazard ratios in randomised controlled trials and general practice research database (GPRD)

	Death	Myocardial infarction	Stroke	CABG/PTCA
Syst-Eur				
Trial	0.86 (0.67 to 1.09)	0.70 (0.44 to 1.09)	0.58 (0.40 to 0.83)	_
GPRD-Cox	1.23 (1.00 to 1.50)*	0.74 (0.52 to 1.07)	0.68 (0.51 to 0.94)	_
WHI-intact uterus				
Trial	0.98 (0.82 to 1.15)	1.11 (0.84 to 1.47)†	1.41 (1.07 to 1.85)	1.01 (0.83 to 1.22)
GPRD-Cox	0.75 (0.65 to 0.86)*	0.95 (0.78 to 1.16)	1.23 (0.99 to 1.52)	1.15 (0.79 to 1.67)
GPRD-PERR	_	1.40 (0.87 to 2.44)	2.63 (1.38 to 7.43)	0.57 (0.22 to 1.56)
GPRD-no missing‡	0.91 (0.79 to 1.05)	_	—	—
WHI-hysterectomy				
Trial	1.01 (0.88 to 1.22)	0.89 (0.70 to 1.12)	1.39 (1.10 to 1.77)	0.93 (0.78 to 1.10)
GPRD-Cox	0.68 (0.57 to 0.81)*	0.50 (0.38 to 0.67)*	0.95 (0.74 to 1.23)*	0.59 (0.36 to 0.95)
GPRD-PERR	_	1.28 (0.69 to 2.56)§	3.06 (1.39 to 10.31)§	1.22 (0.67 to 2.42)
GPRD-no missing‡	0.82 (0.66 to 1.02)	_	—	—
45				
Trial	0.70 (0.58 to 0.85)	0.67 (0.58 to 0.77)	0.64 (0.47 to 0.88)	0.63 (0.54 to 0.74)
GPRD-Cox	0.71 (0.53 to 0.96)	0.79 (0.61 to 1.02)	0.90 (0.63 to 1.30)	2.22 (1.80 to 2.75)*
GPRD-PERR	_	0.69 (0.51 to 0.93)	NA¶	1.00 (0.75 to 1.33)*§
HOPE				
Trial	0.84 (0.75 to 0.95)	0.79 (0.70 to 0.89)	0.68 (0.56 to .84)	0.82 (0.74 to 0.92)
GPRD-Cox	0.94 (0.85 to 1.03)	1.42 (1.23 to 1.61)*	1.16 (0.99 to 1.35)*	1.67 1.34 to 2.07)*
GPRD-PERR	_	0.62 (0.53 to 0.74)*§	0.94 (0.77 to 1.14)*	0.75 (0.56 to 1.01)§
EUROPA				
Trial	0.89 (0.77 to 1.02)	0.76 (0.66 to 0.89)	0.96 (0.72 to 1.28)	0.96 (0.85 to 1.08)
GPRD-Cox	1.06 (0.95 to 1.19)	1.36 (1.16 to 1.58)*	1.04 (0.84 to 1.29)	2.20 (1.85 to 2.62)*
GPRD-PERR		0.84 (0.69 to 1.01)§	0.77 (0.55 to 1.07)	1.26 (.97 to 1.62)§

CABG/PTCA=coronary artery bypass grafts or percutaneous transluminal coronary angioplasty

\*Significant difference (P<0.05) compared with trial.

†Trial values for myocardial infarction reflect WHI re-analysis by age, encompassing 50-70 years.

\$Subset not missing any data for BMI, systolic blood pressure, or smoking.

§Significant difference (P<0.05) compared with GPRD Cox adjusted hazard ratio.

¶PERR could not be done because stroke was study exclusion criteria.

unexposed patients should reflect the aggregate effect of all identified and unidentified confounders.

Therefore, when the unadjusted incidence rate ratio or hazard ratio of that outcome during the study is divided by the ratio for that outcome before the study, this adjustment should correct for the aggregate effects of all identified and unmeasured confounders.

When there are no unmeasured confounders, reflected by similar results of the database Cox adjusted hazard ratio and the randomised controlled trial hazard ratio, the PERR adjusted results should be similar to the Cox adjusted hazard ratio. Based on the empirical findings in these studies, the PERR adjustment seemed to function in this fashion.

When there are unmeasured confounders, presumably resulting from confounding by indication, the results of the PERR adjusted hazard ratio and the Cox adjusted hazard ratio should differ. Our empirical results show that in every instance where the comparison of the Cox adjusted hazard ratio in the database study differed from the results of the trial, suggesting the presence of "unidentified confounding," the PERR adjustment yielded a result much more consistent with the findings in the trial. Of most importance in all but one instance where unmeasured confounding seemed to be present, the PERR adjusted value identified the presence of unmeasured confounding by differing significantly from the Cox adjusted hazard ratio.

Identification of the PERR method emerged from these studies because the direct comparison of the database observational study and the randomised controlled trial provided a presumed correct answer against which to validate the database results. Further investigation is necessary to fully validate the PERR technique. More extensive statistical simulation studies would determine its limitations and applications and the applicability of the method to additional outcomes. It is also important to appreciate that this technique is outcome specific; it cannot be extrapolated from one outcome to another. Finally, it is restricted to outcomes for which previous events can be ascertained. If an outcome was a study exclusion criterion, it cannot be analysed with this approach, nor can it be applied to death.

The PERR method differs and seems to be more widely applicable than other methods that have been developed in an attempt to address hidden bias.<sup>42</sup> As confirmed in our studies, propensity score analysis does not overcome unmeasured confounding. When combined with sensitivity analyses, however, it might provide results that can be interpreted as unlikely to have been influenced by unmeasured covariates.<sup>43,45</sup> Recently, propensity scores combined with regression calibration were used to address unobserved variables under certain conditions.<sup>46 47</sup>

Instrumental variable analysis, used commonly in economics, has also been used to address unmeasured confounding. An instrumental variable analysis requires identification of a factor that affects the assignment to treatment but has no direct effect on the outcome.<sup>48-50</sup> Its applicability and validity for studies of therapeutic efficacy have not been widely examined.<sup>42 5152</sup> Some have suggested that this technique is most suited to address health policy issues rather than specific clinical issues of treatment effectiveness.<sup>48</sup>

Both the propensity score calibration and the instrumental variable analysis methods have important constraints. The propensity score calibration technique requires the presence of a validation study, whereas the instrumental variable analysis requires identification of an appropriate instrument. These requirements limit their applicability to a wide variety of studies.

Of interest, the DID (difference-in-differences) method used in economic studies, has some similarities to the PERR method in that it compares the differences between the difference in before and after behaviour in two groups.<sup>53-55</sup> The key assumption behind the DID method, similar to PERR, is that the distribution of the

Table	3	Hazard ratios a	adiusted for	Cox analysis and	propensity	v scores in general	l practice researd	h database (GPRD)
	- 1				P			

	Death	Myocardial infarction	Stroke	CABG/PTCA
WHI-intact uterus				
Cox	0.75 (0.65 to 0.86)	0.95 (0.78 to 1.16)	1.23 (0.99 to 1.52)	1.15 (0.79 to 1.67)
Propensity score	0.78 (0.69 to 0.89)	0.97 (0.80 to 1.17)	1.16 (0.95 to 1.42)	1.17 (0.82 to 1.70
WHI-hysterectomy				
Cox	0.68 (0.57 to 0.81)	0.50 (0.38 to 0.67)	0.95 (0.74 to 1.23)	0.59 (0.36 to 0.95)
Propensity score	0.67 (0.57 to 0.80)	0.48 (0.36 to 0.64)	0.93 (0.72 to 1.19)	0.59 (0.37 to 0.96)
45				
Cox	0.71 (0.53 to 0.96)	0.79 (0.61 to 1.02)	0.90 (0.63 to 1.30)	2.22 (1.80 to 2.75)
Propensity score	0.65 (0.48 to 0.89)	0.76 (0.59 to 1.00)	0.87 (0.60 to 1.27)	1.93 (1.55 to 2.40)
HOPE				
Cox	0.94 (0.85 to 1.03)	1.42 (1.23 to 1.61)	1.16 (0.99 to 1.35)	1.67 1.34 to 2.07)
Propensity score	0.79 (0.72 to 0.88)*	1.36 (1.17 to 1.58)	1.04 (0.89 to 1.23)	1.55 (1.24 to 1.94)
EUROPA				
Cox	1.06 (0.95 to 1.19)	1.36 (1.16 to 1.58)	1.04 (0.84 to 1.29)	2.20 (1.85 to 2.62)
Propensity score	0.84 (0.75 to 0.95)*	1.31 (1.12 to 1.53)	0.86 (0.69 to 1.08)	2.37 (1.98 to 2.85)

CABG/PTCA=coronary artery bypass grafts or percutaneous transluminal coronary angioplasty \*Significant difference (P<0.05) compared with Cox adjusted hazard ratio.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Two major potential problems could impede the capability of an electronic medical record database to provide reliable information concerning drug efficacy: the quality of the data contained within the database and the ability of analyses of observational—that is, non-experimental—data to provide valid results

The quality of evidence from observational studies is less than from randomised controlled trials because of confounding by indication and other biases related to the effects of unmeasured covariates

## WHAT THIS STUDY ADDS

Although observational studies are subject to unmeasured confounding, a new analytical technique, prior event rate ratio (PERR) adjustment, can identify and reduce unmeasured confounding

Data from properly constructed electronic medical record databases, when analysed with standard statistical methods along with the PERR method, can reveal important insights into the efficacy of medical treatment

unobserved confounding variables in the treated group and the comparison group and the effect of these unobserved confounding variables on the outcome remains the same before and during the study period. The DID method is also used commonly in psychology, where it is called the before and after design with an untreated comparison group.<sup>5657</sup>

Death was significantly higher in one of our database studies (Syst-Eur) and it seemed to be significantly lower in both of the database comparisons with the WHI randomised controlled trial; however, for the reasons enumerated these latter results should be interpreted cautiously.

#### Future perspective and study limitations

Thus it seems from our studies that an electronic medical record database can be an important tool for ascertaining evidence based decisions with regard to treatment. To maximise the value of future databases they should be designed with all the advantages enumerated for GPRD and also should overcome its limitations (see box). Ideally future databases should be much larger than GPRD, which includes about eight million patients. On the basis of our work to date, we estimate that 40-50 million patients are needed for the breadth of future studies we can envisage.

Studies using such databases would not replace the need to do randomised controlled trials but could serve as an important tool to supplement the contributions of trials to evidence based medicine. One example among many is to generalise the results of randomised controlled trials. Although we have not comprehensively examined this issue, our studies have shown the feasibility of further generalising the results of the Syst-Eur and WHI randomised controlled trials.<sup>25 58 59</sup>

As well as the need for further validation of the PERR technique, several other limitations apply to this investigative effort. The PERR technique should be viewed currently as applicable only to analysis of a study using a design similar to ours, which includes similar inclusion and exclusion criteria for the exposed and unexposed and a defined study start, recruitment interval, and end time. Furthermore, the random matching technique might be critical to assure that bias does not exist in the start time for unexposed patients. Application of the PERR technique to other study designs will require its validation under those conditions.

Another potential shortcoming of our studies is the inability to exactly replicate all aspects of the randomised controlled trial independent of randomisation, such as exact dose of study drug, the role of placebos, the possibilities of differences in health care, and other differences between participants entered in randomised controlled trials and those in the general population. In addition, there is also the possibility of inaccuracy of information in the database (for instance, misclassification of outcome, ascertainment bias, etc). The reasonably similar results of the database studies and comparative randomised controlled trials, however, suggest these were not major problems.

Our current view is that the PERR analysis should not be performed in isolation. We would recommend its use along with conventional biostatistical analyses. When the conventional and PERR analyses are similar, "unmeasured confounding" would seem unlikely; whereas when they differ "unmeasured confounding" would seem likely. When unmeasured confounding seems to be present, the PERR analysis seems to yield a more valid result, but additional evaluation is required to ascertain the veracity of this suggestion.

Contributors: RLT and MHW contributed to conception and design; analysis and interpretation of data; drafting and revision of article; and final approval of published version. CX contributed to design, analysis and interpretation of data, drafting and revision of article, and final approval of published version.

We gratefully acknowledge the assistance of Xingmei Wang, who assisted with the biostatistical analyses, and James Lewis and Stephen Kimmel for their insightful review of this manuscript.

**Contributors:** RLT and MHW contributed to conception and design; analysis and interpretation of data; drafting and revision of article; and final approval of published version. CX contributed to design, analysis and interpretation of data, drafting and revision of article, and final approval of published version.

Funding: This work was supported by the National Institutes of Health research grant RO1-HL 073911.

Competing interests: None declared.

Ethical approval: Not required.

Provenance and peer review: Not commissioned; externally peer reviewed.

- Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* 1998:2:1-124.
- 2 MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS. A systematic review of comparisons of effect-sizes derived from randomized and non-randomised studies. *Health Technol Assess* 2000;4:1-154.
- 3 Benson K, Hartz AJ. A comparison of observational studies and randomized controlled trials. *N Engl J Med* 2000;342:1878-86.
- 4 Concato J, Horwitz RI. Beyond randomized versus observational studies. *Lancet* 2004;363:1660-1.
- 5 Sacks HS. Observational studies and randomized trials. *N Engl | Med* 2000;343:1195.
- 6 Ioannidis JPA, Haidich AB, Lau J. Any casualties in the clash of randomised and observational evidence? BMJ 2001;322:879-80.
- 7 Avorn J. In defense of pharmacoepidemiology—embracing the yin and yang of drug research. *N Engl J Med* 2007;357:2219-21.

- 8 Ioannidis JPA, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821-30.
- 9 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
- 10 Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
- 11 Popcock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000;342:1907-9.
- 12 McMahon AD. Observation and experiment with the efficacy of drugs: a warning example from a cohort of nonsteroidal anti-inflammatory and ulcer-healing drug users. *Am J Epidemiol* 2001;154:557-62.
- 13 MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet* 2001;357:455-6.
- 14 Califf RM, Pryor DB, Greenfield J Jr. Beyond randomized clinical trials: applying clinical experience in the treatment of patients with coronary artery disease. *Circulation* 1986;74:1191-4.
- 15 Deeks JJ, Dinnes J, D'Amico RA, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:1-173.
- 16 McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence choosing between randomised and nonrandomised studies. *BMJ* 1999;319:312-5.
- 17 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;342:1887-92.
- 18 Grodstein F, Stampfer MJ, Manson JE, Colditz GA, Willett WC, Rosner B, et al. Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *J Engl J Med* 1996;335:453-61.
- 19 Grodstein F, Manson JE, Colditz GA, Willett WC, Speizer FE, Stampfel MJ. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med* 2000;133:933-41.
- 20 Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in health postmenopausal women. Principal results from the Women's Health Initiative randomized controlled trial. JAMA 2002;288:321-33.
- 21 Vandenbroucke JP. When are observational studies as credible as randomized trials? *Lancet* 2004;363:1728-31.
- 22 Coronary artery surgery study (CASS): a randomized trial of coronary bypass surgery. Comparability of entry characteristics and survival in randomized patients and nonrandomized patients meeting randomization criteria. J Am Coll Cardiol 1984;3:114-28.
- 23 Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. Am J Med 1990;89:630-8
- 24 Gelfand JM, Margolis DJ, Dattani H. The UK general practice research database. In: Strom BL, ed. *Pharmacoepidemiology*. 4th ed. New York: John Wiley, 2005:337-46.
- 25 Tannen RL, Weiner MG, Marcus SM. Simulation of the Syst-Eur randomized control trial using a primary care electronic medical record was feasible. *J Clin Epidemiol* 2006;59:254-64.
- 26 Tannen RL, Weiner MG, Xie D, Barnhart K. A simulation using data from a primary care practice database closely replicated the women's health initiative trial. *J Clin Epidemiol* 2007;60:686-95.
- 27 Tannen RL, Weiner MG, Xie D, Barnhart K. Estrogen affects postmenopausal women differently than estrogen plus progestin therapy. *Hum Reprod* 2007;22:1769-77.
- 28 Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf* 2008;17:661-70.
- 29 Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin converting enzyme inhibitors: further empiric validation of the "prior event rate ratio" to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf* 2008;17:671-85.
- 30 Staessen JA, Fagard R, Thijs L, Celis H, Arabidze GG, Birkenhager WH, et al. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. *Lancet* 1997;350:757-64.
- 31 Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy. The Women's Health Initiative randomized controlled trial. JAMA 2004;291:1701-12.
- 32 Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin survival study (4S). *Lancet* 1994;344:1383-9.

- 33 Heart Outcomes Prevention Evaluation Study Investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high risk patients. *N Engl J Med* 2000;342:145-53.
- 34 European trial on reduction of cardiac events with Perindopril in stable coronary artery disease investigators. Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomized double-blind, placebocontrolled, multicentre trial (the EUROPA study). *Lancet* 2003;362:782-8.
- 35 Raghunathan TE, Solenberger PW, Van Hoewyk J. IVEware: imputation and variance estimation software. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1998.
- 36 Rubin D. *Multiple imputation for non-response in surveys*. New York: John Wiley, 1987.
- 37 Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger PW. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol* 2001;27:85-95.
- 38 Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55.
- 39 Rosenbaum P, Rubin D. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984;79:516-24.
- 40 D'Agostino RB, Rubin DB, Estimating and using propensity scores with partially missing data. J Am Stat Assoc 2000;95:749-59.
- 41 Rossouw JE, Prentice RL, Manson JE, Wu L, Barad D, Barnabei VM, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA* 2007;297:1465-77.
- 42 McMahon AD. Approaches to combat confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003;12:551-8.
- 43 Rosenbaum PR. Discussing hidden bias in observational studies. *Ann Intern Med* 1991;115:901-5.
- 44 Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127:757-63.
- 45 Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. Ann Intern Med 2002;137:693-5.
- 46 Sturmer T, Schneeweiss, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration. A simulation study. *Am J Epidemiol* 2007;165:1110-8.
- 47 Oakes JM, Church TR. Invited commentary: advancing propensity study methods in epidemiology. Am J Epidemiol 2007;165:1119-21.
- 48 McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. JAMA 1994;272:859-66.
- 49 Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health* 1998;19:17-34.
- 50 Sukel TA, Fisher ES, Wennberg DE, Alter DA, Gottleib DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias. Effect of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA* 2007;297:278-85.
- 51 Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables. Applications and limitations. *Epidemiology* 2006;17:260-7.
- 52 Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006;17:268-75.
- 53 Gruber J. The incidence of mandated maternity benefits. *American Economic Review* 1994;84:622-41.
- 54 Meyer BD. Natural and quasi-experiments in economics. *J Business Econom Stat* 13:151-61.
- 55 Angrist JD, Krueger AB. Empirical strategies in labor economics. In: Ashenfelter O, Card D, eds. *Handbook of labor economics*. Amsterdam: Elsevier Science, 1999:1277-366.
- 56 Campbell D, Stanley J. *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton-Mifflin, 1963.
- 57 Shadish WR, Cook TD, Campbell DT. *Experimental and quasiexperimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin, 2002.
- 58 Weiner MG, Barnhart K, Xie D, Tannen RL. Hormone replacement therapy and coronary heart disease in young women. *Menopause* 2008;15:86-93.
- 59 Tannen RL, Weiner MG, Xie D, Barnhart K, Perspectives on hormone replacement therapy: the Women's Health Initiative and new observational studies sampling the overall population. *Fertil Steril* 2008;90:258-64.

Accepted: 19 October 2008

Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application

> U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER) Center for Devices and Radiological Health (CDRH) Center for Food Safety and Applied Nutrition (CFSAN) Center for Veterinary Medicine (CVM) Office of Regulatory Affairs (ORA)

> > August 2003 Pharmaceutical CGMPs

# Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application

Division of Drug Information, HFD-240 Center for Drug Evaluation and Research (CDER) (Tel) 301-827-4573 http://www.fda.gov/cder/guidance/index.htm or Office of Communication, Training and Manufacturers Assistance, HFM-40 Center for Biologics Evaluation and Research (CBER) http://www.fda.gov/cber/guidelines.htm Phone: the Voice Information System at 800-835-4709 or 301-827-1800

or Communications Staff (HFV-12), Center for Veterinary Medicine (CVM) (Tel) 301-594-1755 http://www.fda.gov/cvm/guidance/guidance.html or Division of Small Manufacturers Assistance (HFZ-220) http://www.fda.gov/cdrh/ggpmain.html Manufacturers Assistance Phone Number: 800.638.2041 or 301.443.6597 Internt'l Staff Phone: 301.827.3993 Or

> Center for Food Safety and Applied Nutrition (CFSAN) http://www.cfsan.fda.gov/~dms/guidance.html.

U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER) Center for Devices and Radiological Health (CDRH) Center for Food Safety and Applied Nutrition (CFSAN) Center for Veterinary Medicine (CVM) Office of Regulatory Affairs (ORA)

> August 2003 Pharmaceutical CGMPs

# TABLE OF CONTENTS

I.	INTRODUCTION	.1
II.	BACKGROUND	. 2
III.	DISCUSSION	.3
A.	Overall Approach to Part 11 Requirements	.3
B.	Details of Approach – Scope of Part 11	.4
	1. Narrow Interpretation of Scope	. 4
	2. Definition of Part 11 Records	. 5
C.	Approach to Specific Part 11 Requirements	. 6
	1. Validation	. 6
	2. Audit Trail	. 6
	3. Legacy Systems	. 7
	4. Copies of Records	. 7
	5. Record Retention	. 8
IV.	REFERENCES	.9

# Guidance for Industry<sup>1</sup> Part 11, Electronic Records; Electronic Signatures — Scope and Application

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

# 13 14

15

1

2

8

9

10

11 12

# 16 I. INTRODUCTION 17

18 This guidance is intended to describe the Food and Drug Administration's (FDA's) current

19 thinking regarding the scope and application of part 11 of Title 21 of the Code of Federal

20 Regulations; Electronic Records; Electronic Signatures (21 CFR Part 11).<sup>2</sup> 21

22 This document provides guidance to persons who, in fulfillment of a requirement in a statute or

another part of FDA's regulations to maintain records or submit information to FDA,<sup>3</sup> have

chosen to maintain the records or submit designated information electronically and, as a result,

25 have become subject to part 11. Part 11 applies to records in electronic form that are created,

26 modified, maintained, archived, retrieved, or transmitted under any records requirements set

27 forth in Agency regulations. Part 11 also applies to electronic records submitted to the Agency

under the Federal Food, Drug, and Cosmetic Act (the Act) and the Public Health Service Act (the
PHS Act), even if such records are not specifically identified in Agency regulations (§ 11.1).

The underlying requirements set forth in the Act, PHS Act, and FDA regulations (§ 11.1).

31 11) are referred to in this guidance document as *predicate rules*.

# 32

<sup>2</sup> 62 FR 13430

<sup>&</sup>lt;sup>1</sup> This guidance has been prepared by the Office of Compliance in the Center for Drug Evaluation and Research (CDER) in consultation with the other Agency centers and the Office of Regulatory Affairs at the Food and Drug Administration.

<sup>&</sup>lt;sup>3</sup> These requirements include, for example, certain provisions of the Current Good Manufacturing Practice regulations (21 CFR Part 211), the Quality System regulation (21 CFR Part 820), and the Good Laboratory Practice for Nonclinical Laboratory Studies regulations (21 CFR Part 58).

33 As an outgrowth of its current good manufacturing practice (CGMP) initiative for human and animal drugs and biologics,<sup>4</sup> FDA is re-examining part 11 as it applies to all FDA regulated 34 35 products. We anticipate initiating rulemaking to change part 11 as a result of that re-36 examination. This guidance explains that we will narrowly interpret the scope of part 11. While 37 the re-examination of part 11 is under way, we intend to exercise enforcement discretion with 38 respect to certain part 11 requirements. That is, we do not intend to take enforcement action to 39 enforce compliance with the validation, audit trail, record retention, and record copying 40 requirements of part 11 as explained in this guidance. However, records must still be maintained 41 or submitted in accordance with the underlying predicate rules, and the Agency can take

- 42 regulatory action for noncompliance with such predicate rules.
- 43

44 In addition, we intend to exercise enforcement discretion and do not intend to take (or

45 recommend) action to enforce any part 11 requirements with regard to systems that were

46 operational before August 20, 1997, the effective date of part 11 (commonly known as legacy 47 systems) under the circumstances described in section III C 3 of this guidence

47 systems) under the circumstances described in section III.C.3 of this guidance.

48

49 *Note that part 11 remains in effect* and that this exercise of enforcement discretion applies only
 50 as identified in this guidance.

51

52 FDA's guidance documents, including this guidance, do not establish legally enforceable 53 responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should 54 be viewed only as recommendations, unless specific regulatory or statutory requirements are 55 cited. The use of the word *should* in Agency guidances means that something is suggested or 56 recommended, but not required.

57 58

# 59 II. BACKGROUND

60

In March of 1997, FDA issued final part 11 regulations that provide criteria for acceptance by FDA, under certain circumstances, of electronic records, electronic signatures, and handwritten signatures executed to electronic records as equivalent to paper records and handwritten signatures executed on paper. These regulations, which apply to all FDA program areas, were intended to permit the widest possible use of electronic technology, compatible with FDA's responsibility to protect the public health.

67

68 After part 11 became effective in August 1997, significant discussions ensued among industry,

69 contractors, and the Agency concerning the interpretation and implementation of the regulations.

FDA has (1) spoken about part 11 at many conferences and met numerous times with an industry

71 coalition and other interested parties in an effort to hear more about potential part 11 issues; (2)

72 published a compliance policy guide, CPG 7153.17: Enforcement Policy: 21 CFR Part 11;

73 Electronic Records; Electronic Signatures; and (3) published numerous draft guidance

74 documents including the following:

<sup>&</sup>lt;sup>4</sup> See Pharmaceutical CGMPs for the 21st Century: A Risk-Based Approach; A Science and Risk-Based Approach to Product Quality Regulation Incorporating an Integrated Quality Systems Approach at www.fda.gov/oc/guidance/gmp.html.

75	
76	• 21 CFR Part 11; Electronic Records; Electronic Signatures, Validation
77	• 21 CFR Part 11; Electronic Records; Electronic Signatures, Glossary of Terms
78	• 21 CFR Part 11; Electronic Records; Electronic Signatures, Time Stamps
79	• 21 CFR Part 11; Electronic Records; Electronic Signatures, Maintenance of Electronic
80	Records
81	• 21 CFR Part 11; Electronic Records; Electronic Signatures, Electronic Copies of
82	Electronic Records
83	
84	Throughout all of these communications, concerns have been raised that some interpretations of
85	the part 11 requirements would (1) unnecessarily restrict the use of electronic technology in a
86	manner that is inconsistent with FDA's stated intent in issuing the rule, (2) significantly increase
87	the costs of compliance to an extent that was not contemplated at the time the rule was drafted,
88	and (3) discourage innovation and technological advances without providing a significant public
89	health benefit. These concerns have been raised particularly in the areas of part 11 requirements
90	for validation, audit trails, record retention, record copying, and legacy systems.
91	
92	As a result of these concerns, we decided to review the part 11 documents and related issues,
93	particularly in light of the Agency's CGMP initiative. In the <i>Federal Register</i> of February 4,
94	2003 (68 FR 5645), we announced the withdrawal of the draft guidance for industry, 21 CFR
95	Part 11; Electronic Records; Electronic Signatures, Electronic Copies of Electronic Records.
90	draft guidenee when that draft guidenee may no longer represent our enpresed under the CCMP
97	initiative. Then in the <i>Eaderal Register</i> of February 25, 2003 (68 FP 8775), we appround the
99	withdrawal of the part 11 draft guidance documents on validation glossary of terms time
100	stamps <sup>5</sup> maintenance of electronic records and CPG 7153 17. We received valuable public
101	comments on these draft guidances, and we plan to use that information to help with future
102	decision-making with respect to part 11. We do not intend to re-issue these draft guidance
103	documents or the CPG.
104	
105	We are now re-examining part 11, and we anticipate initiating rulemaking to revise provisions of
106	that regulation. To avoid unnecessary resource expenditures to comply with part 11
107	requirements, we are issuing this guidance to describe how we intend to exercise enforcement
108	discretion with regard to certain part 11 requirements during the re-examination of part 11. As
109	mentioned previously, part 11 remains in effect during this re-examination period.
110	
111	
112	III. DISCUSSION
113	
114	A. Overall Approach to Part 11 Requirements
115	

<sup>&</sup>lt;sup>5</sup> Although we withdrew the draft guidance on time stamps, our current thinking has not changed in that when using time stamps for systems that span different time zones, we do not expect you to record the signer's local time. When using time stamps, they should be implemented with a clear understanding of the time zone reference used. In such instances, system documentation should explain time zone references as well as zone acronyms or other naming conventions.

116 117	As described in more detail below, the approach outlined in this guidance is based on three main elements:
118 119 120	• Part 11 will be interpreted narrowly; we are now clarifying that fewer records will be considered subject to part 11.
121 122 123 124 125	• For those records that remain subject to part 11, we intend to exercise enforcement discretion with regard to part 11 requirements for validation, audit trails, record retention, and record copying in the manner described in this guidance and with regard to all part 11 requirements for systems that were operational before the effective date of part 11 (also known as legacy systems).
126 127	• We will enforce all predicate rule requirements, including predicate rule record and recordkeeping requirements.
128 129 130 131 132 133 134	It is important to note that FDA's exercise of enforcement discretion as described in this guidance is limited to specified part 11 requirements (setting aside legacy systems, as to which the extent of enforcement discretion, under certain circumstances, will be more broad). We intend to enforce all other provisions of part 11 including, but not limited to, certain controls for closed systems in § 11.10. For example, we intend to enforce provisions related to the following controls and requirements:
135	limiting system access to authorized individuals
136	• use of operational system checks
137	• use of authority checks
138	• use of device checks
139	• determination that persons who develop, maintain, or use electronic systems have the
140	education, training, and experience to perform their assigned tasks
141 142	• establishment of and adherence to written policies that hold individuals accountable for actions initiated under their electronic signatures
143	<ul> <li>appropriate controls over systems documentation</li> </ul>
144 145	<ul> <li>controls for open systems corresponding to controls for closed systems bulleted above (§ 11.30)</li> </ul>
146 147 148	• requirements related to electronic signatures (e.g., §§ 11.50, 11.70, 11.100, 11.200, and 11.300)
140	We expect continued compliance with these provisions and we will continue to enforce them
150	Furthermore, persons must comply with applicable predicate rules, and records that are required
151	to be maintained or submitted must remain secure and reliable in accordance with the predicate
152	rules.
153	
154	B. Details of Approach – Scope of Part 11
155	
156	1. Narrow Interpretation of Scope
157	
158 159	We understand that there is some confusion about the scope of part 11. Some have understood the scope of part 11 to be very broad. We believe that some of those broad interpretations could

160 lead to unnecessary controls and costs and could discourage innovation and technological 161 advances without providing added benefit to the public health. As a result, we want to clarify 162 that the Agency intends to interpret the scope of part 11 narrowly. 163 164 Under the narrow interpretation of the scope of part 11, with respect to records required to be 165 maintained under predicate rules or submitted to FDA, when persons choose to use records in 166 electronic format in place of paper format, part 11 would apply. On the other hand, when persons use computers to generate paper printouts of electronic records, and those paper records 167 168 meet all the requirements of the applicable predicate rules and persons rely on the paper records 169 to perform their regulated activities, FDA would generally not consider persons to be "using 170 electronic records in lieu of paper records" under §§ 11.2(a) and 11.2(b). In these instances, the 171 use of computer systems in the generation of paper records would not trigger part 11. 172 173 2. Definition of Part 11 Records 174 175 Under this narrow interpretation, FDA considers part 11 to be applicable to the following records 176 or signatures in electronic format (part 11 records or signatures): 177 178 • Records that are required to be maintained under predicate rule requirements and that are 179 maintained in electronic format *in place of paper format*. On the other hand, records (and 180 any associated signatures) that are not required to be retained under predicate rules, but 181 that are nonetheless maintained in electronic format, are not part 11 records. 182 We recommend that you determine, based on the predicate rules, whether specific records 183 are part 11 records. We recommend that you document such decisions. 184 185 • Records that are required to be maintained under predicate rules, that are maintained in 186 electronic format *in addition to paper format*, and that *are relied on to perform regulated* 187 activities. 188 In some cases, actual business practices may dictate whether you are *using* electronic 189 records instead of paper records under § 11.2(a). For example, if a record is required to 190 be maintained under a predicate rule and you use a computer to generate a paper printout 191 of the electronic records, but you nonetheless rely on the electronic record to perform 192 regulated activities, the Agency may consider you to be *using* the electronic record 193 instead of the paper record. That is, the Agency may take your business practices into 194 account in determining whether part 11 applies. 195 Accordingly, we recommend that, for each record required to be maintained under predicate rules, you determine in advance whether you plan to rely on the electronic 196 197 record or paper record to perform regulated activities. We recommend that you 198 document this decision (e.g., in a Standard Operating Procedure (SOP), or specification 199 document). 200 • Records submitted to FDA, under predicate rules (even if such records are not 201 specifically identified in Agency regulations) in electronic format (assuming the records 202 have been identified in docket number 92S-0251 as the types of submissions the Agency 203 accepts in electronic format). However, a record that is not itself submitted, but is used

204 205	in generating a submission, is not a part 11 record unless it is otherwise required to be maintained under a predicate rule and it is maintained in electronic format.			
206 207 208 209 210 211	• Electronic signatures that are intended to be the equivalent of handwritten signatures, initials, and other general signings required by predicate rules. Part 11 signatures include electronic signatures that are used, for example, to document the fact that certain events or actions occurred in accordance with the predicate rule (e.g. <i>approved</i> , <i>reviewed</i> , and <i>verified</i> ).			
212	C. Approach to Specific Part 11 Requirements			
213				
214	1. Validation			
215				
216	The Agency intends to exercise enforcement discretion regarding specific part 11 requirements for validation of computarized systems ( $\$$ 11 10(a) and corresponding requirements in $\$$ 11 20)			
217	Although persons must still comply with all applicable predicate rule requirements for validation			
218	$(e_{g} 21 \text{ CFR } 820.70(i))$ this guidance should not be read to impose any additional requirements			
220	for validation			
220				
222	We suggest that your decision to validate computerized systems, and the extent of the validation,			
223	take into account the impact the systems have on your ability to meet predicate rule			
224	requirements. You should also consider the impact those systems might have on the accuracy,			
225	reliability, integrity, availability, and authenticity of required records and signatures. Even if			
226	there is no predicate rule requirement to validate a system, in some instances it may still be			
227	important to validate the system.			
228				
229	We recommend that you base your approach on a justified and documented risk assessment and			
230	a determination of the potential of the system to affect product quality and safety, and record			
231	integrity. For instance, validation would not be important for a word processor used only to			
232	generate SOPs.			
233				
234	For further guidance on validation of computerized systems, see FDA's guidance for industry			
235	and FDA staff General Principles of Software Vallaation and also industry guidance such as the			
230	GAMF 4 Guide (See References).			
237	2 Audit Trail			
230				
240	The Agency intends to exercise enforcement discretion regarding specific part 11 requirements			
241	related to computer-generated, time-stamped audit trails (§ 11.10 (e), (k)(2) and any			
242	corresponding requirement in §11.30). Persons must still comply with all applicable predicate			
243	rule requirements related to documentation of, for example, date (e.g., § 58.130(e)), time, or			
244	sequencing of events, as well as any requirements for ensuring that changes to records do not			
245	obscure previous entries.			
246				
247	Even if there are no predicate rule requirements to document, for example, date, time, or			
248	sequence of events in a particular instance, it may nonetheless be important to have audit trails or			

other physical, logical, or procedural security measures in place to ensure the trustworthiness and

250	reliability of the records. <sup>6</sup> We recommend that you base your decision on whether to apply audit				
251	trails, or other appropriate measures, on the need to comply with predicate rule requirements, a				
252	justified and documented risk assessment, and a determination of the potential effect on product				
253	quality and safety and record integrity. We suggest that you apply appropriate controls based on				
254	such an assessment. Audit trails can be particularly appropriate when users are expected to				
255	create, modify, or delete regulated records during normal operation.				
256					
257	3. Legacy Systems <sup>7</sup>				
258					
259	The Agency intends to exercise enforcement discretion with respect to all part 11 requirements				
260	for systems that otherwise were operational prior to August 20, 1997, the effective date of part				
261	11, under the circumstances specified below.				
262					
263	This means that the Agency does not intend to take enforcement action to enforce compliance				
264	with any part 11 requirements if all the following criteria are met for a specific system:				
265					
266	• The system was operational before the effective date.				
267	• The system met all applicable predicate rule requirements before the effective date.				
268	<ul> <li>The system currently meets all applicable predicate rule requirements</li> </ul>				
260	<ul> <li>The system currently meets an applicable predicate full requirements.</li> <li>You have documented evidence and justification that the system is fit for its intended use.</li> </ul>				
20)	(including having an acceptable level of record security and integrity, if applicable)				
270	(including having an acceptable lever of record security and integrity, if applicable).				
271	If a system has been changed since August 20, 1997, and if the changes would prevent the				
272	system from meeting predicate rule requirements. Part 11 controls should be applied to Part 11				
273	records and signatures pursuant to the enforcement policy expressed in this guidance				
275	records and signatures pursuant to the enforcement policy expressed in this guidance.				
275	4 Copies of Records				
270	4. Copies of Records				
278	The Agency intends to exercise enforcement discretion with regard to specific part 11				
279	requirements for generating copies of records (8 11 10 (b) and any corresponding requirement in				
280	811.30) You should provide an investigator with reasonable and useful access to records during				
281	an inspection All records held by you are subject to inspection in accordance with predicate				
282	rules (e.g. $\&$ 211 180(c) (d) and 108 35(c)(3)(ii))				
282	Tules (e.g., §§ 211.100(c), (d), and 100.55(c)(5)(1)).				
285	We recommend that you supply copies of electronic records by:				
285	we recommend that you supply copies of electronic records by.				
205	• Producing copies of records held in common portable formats when records are				
280	• I foldering copies of fectorus nerd in common portable formats when fectorus are				
201	manualited in these formats				
288	• Using established automated conversion or export methods, where available, to make				
289	copies in a more common format (examples of such formats include, but are not limited				
290	to, PDF, XML, or SGML)				

<sup>&</sup>lt;sup>6</sup> Various guidance documents on information security are available (see References).

<sup>&</sup>lt;sup>7</sup> In this guidance document, we use the term *legacy system* to describe systems already in operation before the effective date of part 11.

291 In each case, we recommend that the copying process used produces copies that preserve the 292 content and meaning of the record. If you have the ability to search, sort, or trend part 11 293 records, copies given to the Agency should provide the same capability if it is reasonable and 294 technically feasible. You should allow inspection, review, and copying of records in a human 295 readable form at your site using your hardware and following your established procedures and 296 techniques for accessing records.

297 298

5. **Record Retention** 

299

300 The Agency intends to exercise enforcement discretion with regard to the part 11 requirements 301 for the protection of records to enable their accurate and ready retrieval throughout the records 302 retention period (§ 11.10 (c) and any corresponding requirement in §11.30). Persons must still 303 comply with all applicable predicate rule requirements for record retention and availability (e.g., 304 §§ 211.180(c),(d), 108.25(g), and 108.35(h)).

305

306 We suggest that your decision on how to maintain records be based on predicate rule

307 requirements and that you base your decision on a justified and documented risk assessment and 308 a determination of the value of the records over time.

309

310 FDA does not intend to object if you decide to archive required records in electronic format to

311 nonelectronic media such as microfilm, microfiche, and paper, or to a standard electronic file

312 format (examples of such formats include, but are not limited to, PDF, XML, or SGML).

313 Persons must still comply with all predicate rule requirements, and the records themselves and

314 any copies of the required records should preserve their content and meaning. As long as

315 predicate rule requirements are fully satisfied and the content and meaning of the records are

preserved and archived, you can delete the electronic version of the records. In addition, paper 316

317 and electronic record and signature components can co-exist (i.e., a hybrid<sup>8</sup> situation) as long as

318 predicate rule requirements are met and the content and meaning of those records are preserved.

319

<sup>&</sup>lt;sup>8</sup> Examples of hybrid situations include combinations of paper records (or other nonelectronic media) and electronic records, paper records and electronic signatures, or handwritten signatures executed to electronic records.

319		
320	IV.	REFERENCES

320	IV.	REFERENCES
321		
322	Fo	od and Drug Administration References
323		
324	1.	Glossary of Computerized System and Software Development Terminology (Division of
325		Field Investigations, Office of Regional Operations, Office of Regulatory Affairs, FDA
326		1995) (http://www.fda.gov/ora/inspect_ref/igs/gloss.html)
327		
328	2.	General Principles of Software Validation; Final Guidance for Industry and FDA Staff
329		(FDA, Center for Devices and Radiological Health, Center for Biologics Evaluation and
330		Research, 2002) (http://www.fda.gov/cdrh/comp/guidance/938.html)
331		
332	3.	Guidance for Industry, FDA Reviewers, and Compliance on Off-The-Shelf Software Use
333		in Medical Devices (FDA, Center for Devices and Radiological Health, 1999)
334		(http://www.fda.gov/cdrh/ode/guidance/585.html)
335		
336	4.	Pharmaceutical CGMPs for the 21 <sup>st</sup> Century: A Risk-Based Approach; A Science and
337		Risk-Based Approach to Product Quality Regulation Incorporating an Integrated Quality
338		Systems Approach (FDA 2002) (http://www.fda.gov/oc/guidance/gmp.html)
339		
340	_	
341	In	dustry References
342		
343	1.	The Good Automated Manufacturing Practice (GAMP) Guide for Validation of
344		Automated Systems, GAMP 4 (ISPE/GAMP Forum, 2001) (http://www.ispe.org/gamp/)
345	•	
346	2.	ISO/IEC 17/99:2000 (BS 7/99:2000) Information technology – Code of practice for
347		information security management (ISO/IEC, 2000)
348	2	
349 250	3.	ISO 149/1:2002 Medical Devices- Application of risk management to medical devices
350		(150, 2001)
331 252		
552		

[Code of Federal Regulations]
[Title 21, Volume 1]
[Revised as of April 1, 2011]
[CITE: 21CFR11.100]

## TITLE 21--FOOD AND DRUGS CHAPTER I--FOOD AND DRUG ADMINISTRATION DEPARTMENT OF HEALTH AND HUMAN SERVICES SUBCHAPTER A--GENERAL

PART 11 -- ELECTRONIC RECORDS; ELECTRONIC SIGNATURES

Subpart C--Electronic Signatures

Sec. 11.100 General requirements.

(a) Each electronic signature shall be unique to one individual and shall not be reused by, or reassigned to, anyone else.

(b) Before an organization establishes, assigns, certifies, or otherwise sanctions an individual's electronic signature, or any element of such electronic signature, the organization shall verify the identity of the individual.

(c) Persons using electronic signatures shall, prior to or at the time of such use, certify to the agency that the electronic signatures in their system, used on or after August 20, 1997, are intended to be the legally binding equivalent of traditional handwritten signatures.

(1) The certification shall be submitted in paper form and signed with a traditional handwritten signature, to the Office of Regional Operations, 12420 Parklawn Drive, RM 3007 Rockville, MD 20857.

# Data Analysis Projects

This is an initial list of US projects to analyze and learn from clinical data put together by Roundtable staff. It is not an exhaustive list, and inclusion does not denote endorsement. We welcome your input on additional efforts that could be added.

Name	Major Collaborators	Focus
Analgesic Clinical Trials Innovation, Opportunities, and Networks (ACTION) Initiative	University of Rochester University of Pennsylvania FDA	Data analyses of analgesic clinical trial data to determine the effects of specific research designs and analysis methods
Cardiac Safety Research Consortium (CSRC)	Duke FDA Industry partnership	To advance scientific knowledge on cardiac safety for new and existing medical products
Coalition Against Major Diseases (CAMD)	Critical Path Institute FDA Patient organizations Medical product industry Brookings Institution	Development of a disease progression model for Alzheimer's disease
Healthcare Cost and Utilization Project (H- CUP)	Federal-State-Industry partnership AHRQ	Largest collection of longitudinal hospital care data in the United States
Health Care Cost Institute	Aetna Humana Kaiser Permanente UnitedHealthcare	Comprehensive source of information on health care costs and utilization Promoting research on the drivers of escalating health care costs and utilization
HealthData.gov	National Cancer Institute American Association of Diabetes Educators AT&T Baylor University Institute of Medicine	Development of electronic applications to promote health care
High Value Healthcare Collaborative	Mayo Clinic Denver Health Intermountain Healthcare Dartmouth-Hitchcock Cleveland Clinic Dartmouth Institute for Health Policy and Clinical Practice Baylor Health Care System Beaumont Hospitals	Research in nine increasingly prevalent condition/disease-specific areas that have been shown to have wide variation in rates, costs, and outcomes

	MaineHealth University of Iowa Health Care Scott & White Health Care Sutter Health UCLA Health System Virginia Mason Medical Center	
HMO Research Network	Consortium of 19 health care delivery organizations	Research areas include cardiovascular, diabetes, pregnancy, mental health, therapeutics, safety of medical products, Therapeutic effectiveness and Information Technology
Multi-Payer Claims Database (MPCD) for Comparative Effectiveness Research	HHS ASPE CMS	Consolidating access to longitudinal data on health services financed by both public and private payers to help facilitate CER
Observational Medical Outcomes Partnership	Pharmaceutical industry Academic institutions Non-profit organizations FDA Other federal agencies	Monitoring of drugs for safety
Predictive Safety Testing Consortium (PSTC)	16 pharmaceutical companies FDA European Medicines Agency	Conducting prospective studies to generate biomarker qualification packages for evaluation by the FDA.
Query Health	ONC	To establish standards and services for distributed population health queries
Stanford Translational Research Integrated Database Environment (STRIDE):	Stanford University	To create a standards-based informatics platform supporting clinical and translational research
THEW (University of Rochester Telemetric and Holter ECG Warehouse)	FDA University of Rochester Other public and private stakeholders.	To develop automatic ECG analysis algorithms.

# Data Analysis Projects

# Analgesic Clinical Trials Innovation, Opportunities, and Networks (ACTION) Initiative

- The ACTION Initiative Public-Private Partnership (PPP) is being implemented by the University of Rochester's Center for Human Experimental Therapeutics in cooperation with the U.S. Food and Drug Administration (FDA).
- Partners with University of Pennsylvania and other societies
- Conducting in-depth and wide-ranging data analyses of analgesic clinical trial data to determine the effects of specific research designs and analysis methods.

# Source:

http://www.fda.gov/AboutFDA/PartnershipsCollaborations/PublicPrivatePartn ershipProgram/ucm231130.htm

# Cardiac Safety Research Consortium (CSRC):

- Duke/FDA/Industry partnership
- To advance scientific knowledge on cardiac safety for new and existing medical products
- Projects:
  - Safety of Atrial Fibrillation Ablation Initiative (SAFARI)
  - Dual Anti-Platelet Therapy (DAPT) Study
  - CSRC-HESI (Health and Environmental Sciences Institute) preclinical cardiac safety evaluation
  - Ten "Thorough QT" studies (Placebo and Moxifloxacin control arms only) released from FDA ECG warehouse by data owners (Merck, GSK, Lilly) for use in CSRC-approved research projects. Four studies are currently available accounting for a total of 25,000 individual records.
  - CSRC TransRadial Education And Training (TREAT)
  - CSRC Cardiac Troponin evaluation

# Source:

http://www.fda.gov/AboutFDA/PartnershipsCollaborations/PublicPrivatePartners hipProgram/ucm231121.htm

# Coalition Against Major Diseases (CAMD)

- CAMD database of 4000 patients with Alzheimer's disease
- CAMD was formed by the non-profit Critical Path Institute (C-Path), in cooperation with the FDA, patient organizations, the medical products

industry and the Engelberg Center for Health Care Reform at the Brookings Institution.

• Development of a disease progression model for Alzheimer's disease using data from the NIH-sponsored Alzheimer's Disease Neuroimaging Initiative (ADNI) database. A briefing package was submitted to the FDA in December 2009 for formal review. Feedback was provided by FDA in April of 2010 and has been incorporated into the model.

Source:

http://www.fda.gov/AboutFDA/PartnershipsCollaborations/PublicPrivatePartners hipProgram/ucm231134.htm

# Healthcare Cost and Utilization Project (H-CUP):

- Family of health care databases developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ)
- Largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988 including:

**The Nationwide Inpatient Sample (NIS)** with inpatient data from a national sample of over 1,000 hospitals.

**The Kids' Inpatient Database (KID)** is a nationwide sample of pediatric inpatient discharges.

**The Nationwide Emergency Department Sample (NEDS)** is a database that yields national estimates of emergency department (ED) visits.

**The State Inpatient Databases (SID)** contain the universe of inpatient discharge abstracts from participating states.

**The State Ambulatory Surgery Databases (SASD)** contain data from ambulatory care encounters from hospital-affiliated and sometimes freestanding ambulatory surgery sites.

**The State Emergency Department Databases (SEDD)** contain data from hospital-affiliated emergency departments for visits that do not result in hospitalizations.

# • Objectives:

Create and enhance a powerful source of national, state, and all-payer health care data. Produce a broad set of software tools and products to facilitate the use of HCUP and other administrative data. Enrich a collaborative partnership with statewide data organizations aimed at increasing the quality and use of health care data. Conduct and translate research to inform decision making and improve health care delivery.

# Source: http://www.hcup-us.ahrq.gov/overview.jsp

# **Health Care Cost Institute**

- A new non-profit research institute and a unique and unprecedented health research partnership to promote independent research and analysis on the causes of rising US health spending, to provide policy makers, consumers, and researchers with better, more transparent information on what is driving health care costs
- Several major health insurers will provide information on billions of medicalbilling claims in their books to the new academic institute, which will create a database for research on health-care costs and utilization
- The database includes 5,000 hospitals and more than 1 million different medical service providers from commercial health plans operated by Aetna, Humana, Kaiser Permanente and UnitedHealthcare and the Medicare Advantage data from each participating plan

# Source: <a href="http://healthcostinstitute.org/">http://healthcostinstitute.org/</a>

# HealthData.gov:

• High-value datasets, tools, and applications using data about health and healthcare

**mHealth Initiative:** HHS formed the Text4Health Task Force which wrote recommendations to guide HHS' strategic development of health text messaging and accelerate the growth of mHealth and health innovation.

The National Cancer Institute (NCI) at the National Institutes of Health is launching the SmokeFreeTXT program

HHS Office of Minority Health has launched a collaborative effort in partnership with American Association of Diabetes Educators (AADE), AT&T, and Baylor University to investigate the use of smart phones' secure video streaming by demonstrating live clinician/community health worker directed diabetes selfmanagement education courses

HHS has partnered with the White House to launch the Apps Against Abuse developer's challenge

**Health Data Initiative:** HHS and the Institute of Medicine have launched a national initiative to help consumers and communities get more value out of the Nation's wealth of health data through the development of applications for use by providers and the public.

Source: http://www.hhs.gov/open/initiatives/index.html

# **High Value Healthcare Collaborative**

- Eight major health systems join the Collaborative including Mayo Clinic, Denver Health, Intermountain Healthcare, Dartmouth-Hitchcock, Cleveland Clinic, The Dartmouth Institute for Health Policy and Clinical Practice (TDI), Baylor Health Care System, Beaumont Hospitals, MaineHealth, Scott & White Health Care, Sutter Health, UCLA Health System, University of Iowa Health Care, and Virginia Mason Medical Center to improve health care, lower costs, and move best practices out to the national provider community.
- Working together in nine increasingly prevalent condition/disease-specific areas that have been shown to have wide variation in rates, costs, and outcomes nationally including Total Knee Replacement; Diabetes; Asthma; Hip Surgery; Heart Failure; Perinatal Care; Depression; Spine Surgery and Weight Loss Surgery

# Source: <a href="http://www.dartmouth-hitchcock.org/news/newsdetail/59724/">http://www.dartmouth-hitchcock.org/news/newsdetail/59724/</a>

# **HMO Research Network**

- The HMO Research Network is a consortium of 19 health care delivery organizations with both defined patient populations and formal, recognized research capabilities.
- Projects:
- 1. **Cancer Research Network (CRN):** addresses the spectrum of cancer control, including prevention, early detection, treatment, survivorship, surveillance, and end-of-life care through a program of collaborative research for cancer prevention and control.
- 2. **Cardiovascular Research Network (CVRN):** a framework to answer critical questions about contemporary cardiovascular epidemiology, optimal management, and associated clinical outcomes and resource utilization within large community-based populations
- 3. **Center for Education and Research on Therapeutics (CERT):** to conduct research and provide education that will advance the optimal use of drugs, medical devices, and biological products; increase awareness of the benefits and risks of therapeutics
- 4. **Developing Evidence to Improve Decisions about Effectiveness** (**DEcIDE**) **Network:** The HMORN DEcIDE-1 Network is made of twelve of the HMORN research centers uses using the health plans' defined populations, providers, delivery systems, and unique data resources to develop information about therapeutic effectiveness within typical clinical settings. The HMORN DEcIDE-2 Network is made up of 14 HMORN research centers and develops scientific evidence and methodologies about the outcomes, comparative clinical effectiveness, safety, and appropriateness of health care items and services

- 5. **Medical Exposure in Pregnancy Risk Evaluation Program (MEPREP):** The program is a collaboration between the FDA, 11 of the HMORN CERT sites and Vanderbilt University in developing infrastructure to enhance the ability to study the effects of medication exposure during pregnancy on the fetus by creating and maintaining linked data on mothers and infants. One of the latest MEPREP studies to be funded will address the safety of Sulfonamide use by mothers during pregnancy
- 6. **Mental Health Research Network (MHRN) :** Ten HMORN research centers currently participate in the Network in the development of a core infrastructure for collaborative effectiveness research in Mental Health and completion of 4 research projects that leverage that infrastructure in specific clinical areas
- 7. **Mini-Sentinel:** The Mini-Sentinel Coordinating Center is a collaborative effort is consisting of 33 partnering organizations, including 13 HMORN member sites in the development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products using automated healthcare data from multiple sources, a pilot project conducted under contract with the U.S. Food and Drug Administration (FDA)
- 8. **Multi-Institutional Consortium for Comparative Effectiveness Research in Prevention and Treatment of Diabetes Mellitus (SUPREME-DM):** creating comprehensive, standardized diabetes DataLink that contains 1.3 million insured patients with Diabetes Mellitus to identify and monitor trends in diabetes incidence and prevalence, and in diabetes treatment patterns and outcomes.
- 9. Scalable Partnering Network for Comparative Effectiveness Research (SPAN): involves 10 HMORN sites, as well as Denver Health and Hospital Authority in the development of Information Technology (IT), a distributed research network that will link all participating sites through IT and comparative effectiveness research related to obesity and ADHD.
- 10. **Vaccine Safety Datalink (VSD):** to monitor immunization safety and address gaps in scientific knowledge about rare and serious events following immunization on nearly 9.5 million HMORN-based patients annually (more than 3% of the United States population) involving the Centers for Disease Control and Prevention (CDC) America's Health Insurance Plans (AHIP) and today includes ten HMORN member organizations.

Source: <u>http://www.hmoresearchnetwork.org/projects.htm#actionII</u>

# Multi-Payer Claims Database (MPCD) for Comparative Effectiveness Research

• The Multi-Payer Claims Database (MPCD) project is one of a number of initiatives related to comparative effectiveness research (CER) funded by

the American Recovery and Reinvestment Act of 2009. The Act provided \$1.1 billion to build the necessary infrastructure and capacity to support CER.

- Within HHS, the Office of the Assistance Secretary for Planning and Evaluation (ASPE) was tasked with managing the MPCD project in partnership with the Center for Medicare and Medicaid Services (CMS).
- The project represents a private/public partnership with the goal of consolidating access to longitudinal data on health services financed by both public and private payers to help facilitate CER
- the MPCD will initially include claims data, since these data are most readily available. Over time, data with additional clinical detail from other sources, such as EHRs, may be incorporated into the database.

# Source: http://gablog.commpartners.com/?p=309

# **Observational Medical Outcomes Partnership**

- Is a public-private partnership designed to help improve the monitoring of drugs for safety by analyzing existing healthcare databases to identify and evaluate safety and benefit issues of drugs already on the market.
- Partnering with pharmaceutical industry, academic institutions, non-profit organizations, the FDA, and other federal agencies. It is funded and managed through the Foundation for the National Institutes of Health.

Source: <a href="http://omop.fnih.org/">http://omop.fnih.org/</a>

# Predictive Safety Testing Consortium (PSTC)

- The consortium is comprised of sixteen pharmaceutical companies that have signed a legal agreement committing to share their pre-competitive safety testing methods, data and knowledge to advance safety assessments in medical product development. The FDA and the EMA are active advisors and participants in the consortium.
- Sharing preclinical and clinical data for genomic, proteomic and metabolomic biomarkers of drug-induced nephrotoxicity, hepatotoxicity, vascular injury, and carcinogenicity for evaluation and comparison by members of the consortium.
- Conducting prospective studies to generate biomarker qualification packages for evaluation by the FDA.
- Results:

- 1. Critical Path Institute worked with the FDA to pilot the "qualification" process for evaluation of new biomarkers for a specific context of use.
- 2. In 2008, the first sets of biomarkers were submitted to the FDA and EMA by PSTC and were qualified by for use in testing for renal safety in rodents to detect drug-induced kidney injury in laboratory animals.
- 3. The Hepatotoxicity Working Group of PSTC submitted a briefing package to FDA for qualification of four biomarkers for detecting drug-induced liver toxicity.
- 4. The Nephrotoxicity Working Group of PSTC is evaluating the expanded clinical utility of the previously qualified biomarkers and others in human clinical research sponsored by the Foundation for the NIH.
- 5. The Myopathy Working Group has submitted a briefing package of data that support eight novel biomarkers for detecting and monitoring drug-induced skeletal muscle injury in the rat.
- 6. A PSTC database has been established. A common lexicon, study design elements, and standardization of tissue and sample handling, as well as data reporting have been established to facilitate combining the data from multiple studies performed at multiple sites.

# Source:

http://www.fda.gov/AboutFDA/PartnershipsCollaborations/PublicPrivatePartn ershipProgram/ucm231132.htm

# **Query Health:**

- An initiative launched by the Office of the National Coordinator for HIT to establish standards and services for distributed population health queries
- Used to send questions to clinical data sources which return aggregate measures of population health that can be used for many purposes including disease outbreak monitoring, post-market surveillance, comparative effectiveness research, quality and performance measures

Source: http://www.healthdatamanagement.com/news/query-health-distributedqueries-data-analytics-analysis-44013-1.html

# Stanford Translational Research Integrated Database Environment (STRIDE):

- Research and development project at Stanford University to create a standards-based informatics platform supporting clinical and translational research.
- STRIDE consists of three integrated components: a **clinical data warehouse**, based on the HL7 Reference Information Model (RIM), containing clinical

information on over 1.6 million pediatric and adult patients cared for at Stanford University Medical Center since 1995; an application development framework for building research data management applications on the STRIDE-DM platform and a biospecimen data management system.

Source: <u>https://clinicalinformatics.stanford.edu/research/stride.html</u>

**THEW (**University of Rochester Telemetric and Holter ECG Warehouse)

- Implements joint projects among FDA, University of Rochester, and other public and private stakeholders.
- Designed to inform cardiac safety and medical product development, the THEW enables access to unique data and tools to develop automatic ECG analysis algorithms.
- Current research projects aim at increasing QT measurement accuracy used in drug safety trials, which would result in smaller, faster, and therefore more cost- and time-effective drug development processes. Better ECG markers and high precision detection techniques will benefit all stakeholders by ensuring that unsafe drugs do not get to market, while reducing the chances of a "false positive" signal.

## Source:

http://www.fda.gov/AboutFDA/PartnershipsCollaborations/PublicPrivatePartners hipProgram/ucm231128.htm **Biographies and Meeting Logistics** 



# **Speaker Biographies**

James W. Buehler, MD is the Director of the Public Health Surveillance & Informatics Program Office (proposed) at CDC. Dr. Buehler has >30 years of experience in the field of medical epidemiology, serving from 1981-2002 as a Commissioned Officer in the United States Public Health Service at CDC, where he worked in the areas of general field epidemiology, maternal and child health, HIV/AIDS, and, for a brief period in 2001, anthrax. In 2002, Dr. Buehler joined the Epidemiology Department of the Rollins School of Public Health at Emory University, where he held the position of Research Professor. In 2009, he returned to CDC to contribute to the surveillance of pandemic influenza, and in 2010, he became the founding director of CDC's Public Health Surveillance Program Office. Dr. Buehler has devoted much of his career to the field of public health surveillance. As a member of the Emory faculty, Dr. Buehler's research interests centered on improving public health surveillance and emergency preparedness capacities and on advancing the relatively new field of public health systems research. While at Emory, he served as a consultant to epidemiology and emergency preparedness programs at the Division of Public Health of the Georgia Department of Human Resources. In 2006-2008, he served as the public health representative on the Georgia Health Information Technology and Transparency Advisory Board, where he focused on strengthening linkages between public health and healthcare through advances in health information technologies. Dr. Buehler obtained his Bachelor's degree in Biochemistry from the University of California, Berkeley, and his Doctor of Medicine degree from the University of California, San Francisco. He completed residency training in Pediatrics at the University of Oregon Health Sciences Center in Portland and in Preventive Medicine at CDC. He is a Fellow of the American Academy of Pediatrics and is board certified in Pediatrics and Preventive Medicine.

Christopher G. Chute, MD, DrPH received his undergraduate and medical training at Brown University, internal medicine residency at Dartmouth, and doctoral training in Epidemiology at Harvard. He is Board Certified in Internal Medicine, and a Fellow of the American College of Physicians, the American College of Epidemiology, and the American College of Medical Informatics. He became founding Chair of Biomedical Informatics at Mayo in 1988, stepping down after 20 years in that role. He is now Professor of Medical Informatics, and is PI on a large portfolio of research including the HHS/Office of the National Coordinator (ONC) SHARP (Strategic Health IT Advanced Research Projects) on Secondary EHR Data Use, the ONC Beacon Community (Co-PI), the LexGrid projects, Mayo's CTSA Informatics, Mayo's Cancer Center Informatics including caBIG, and several NIH grants including one of the eMERGE centers from NGHRI. Dr. Chute serves as Vice Chair of the Mayo Clinic Data Governance for Health Information Technology Standards, and on Mayo's enterprise IT Oversight Committee. He is presently Chair, ISO Health Informatics Technical Committee (ISO TC215) and Chairs the World Health Organization (WHO) ICD-11 Revision. He also serves on the Health Information Technology Standards Committee for the Office of the National Coordinator in the US DHHS, and the HL7 Advisory Board. Recently held positions include Chair of the Biomedical Computing and Health Informatics study section at NIH, Chair of the Board of the HL7/FDA/NCI/CDISC BRIDG project, on the Board of the Clinical Data Interchange Standards Consortium (CDISC), ANSI Health Information Standards Technology Panel (HITSP) Board member, Chair of the US delegation to ISO TC215 for Health Informatics, Convener of Healthcare Concept Representation WG3 within the (TC215), Co-chair of the HL7 Vocabulary Committee, Chair of the International Medical Informatics Association (IMIA) WG6 on Medical Concept Representation, American Medical Informatics Association (AMIA) Board member, and multiple other NIH biomedical informatics study sections as chair or member.

**Rich Elmore** is ONC's leader for Query Health, an ONC-sponsored initiative to establish standards and services for distributed population queries of electronic health records. He is on a leave of absence from health care technology provider Allscripts, where as Vice President Strategic Initiatives, he managed exploration and execution of acquisitions and strategic partnerships and prior to that he ran the Allscripts Provider Analytics business. He had a long career at IDX where he ran the Flowcast Hospital business and prior to that was Vice President of Product Development for IDX Flowcast. Mr. Elmore was the Communications Workgroup leader for the ONC's Direct Project. He was a charter member of the Interoperability Workgroup for the Certification Commission for Healthcare Information Technology. Mr. Elmore has degrees from Dartmouth College (BA) and New School University (MA Economics). He is on the Board of Directors for Patient Engagement Systems, a chronic disease technology company, and serves as Vice Chair on the Board of Directors for the King Street Center serving kids in need and their families in Burlington Vermont.

**Doug Fridsma, MD, PhD** is the director of the Office of Standards and Interoperability and the Acting Chief Scientist in the Office of the National Coordinator for Health Information Technology. Prior to arriving at ONC, Dr. Fridsma was on the teaching staff in the Department of Biomedical Informatics at Arizona State University and had a clinical practice at Mayo Clinic Scottsdale. Dr. Fridsma completed his medical training at the University of Michigan in 1990, and his PhD in Biomedical Informatics from Stanford University in 2003. In his role at ONC, Dr. Fridsma is responsible for the Nationwide Health Information Network, the Federal Health Architecture, the EHR Certification programs, and other initiatives focused on promoting interoperable health information exchange. He served on the Clinical Data Interchange Standards Consortium (CDISC) Board of Directors from 2005-2008, and was appointed to the Health IT Standards Committee in 2009. He resigned from the HIT Standards Committee after he joined ONC, and recently became a board member of HL7.

James Allen Heywood is the Co-Founder and Chairman of PatientsLikeMe. An MIT engineer, Jamie Heywood entered the field of translational research and medicine when his brother Stephen was diagnosed with ALS in 1998 at the age of 29. With experience in design, information technology, systems modeling, neuroscience and industrial engineering, Heywood brings a unique perspective to drug discovery and medicine. The scientific and business innovations he developed at ALS TDI and PatientsLikeMe have been transforming the intersection of biotechnology and pharmaceutical development, personalized medicine, and patient care. Heywood is the chairman of PatientsLikeMe, where he provides the scientific vision and architecture for its patient-centered medical platform. He cofounded the company in 2005 with his youngest brother, Benjamin, and friend, Jeff Cole. Named one of "15 companies that will change the world" by CNNMoney, PatientsLikeMe is a personalized research and peer care platform that allows patients to share in-depth information on treatments, symptoms and outcomes. This novel open model allows clinicians, providers, and the pharmaceutical industry to better understand diseases and the patient experience. Patients improve their care and actively partner with industry to accelerate and influence the development of new treatments and biomarkers. In 1999 shortly after Stephen was diagnosed, Heywood founded the ALS Therapy Development Institute (ALS TDI), the world's first non-profit biotechnology company, where he served as CEO until 2007. Pioneering an open research model and an industrialized therapeutic validation process, Heywood led ALS TDI to become the world's largest and most comprehensive ALS research program. The comprehensive in-vivo validation program Heywood developed was unable to replicate any of the published preclinical studies of the field that led to human trials calling into question the standards that allowed many drugs to be tested on patients. In 2009, Heywood and a small group of thought leaders founded HealthDataRights.org, an organization that asserts a new patient right to access a copy of all of their medical data in a computable form. Heywood is a published author, frequent speaker, media pundit and an active investment advisor. He speaks at business, government and academic, conferences around the world, including TEDMED, the Milken Global Conference, Health 2.0, Gov 2.0, Personal Democracy

Forum, Institute of Medicine, and the NIH. Heywood is a member of the CDC's National Biosurveillance Advisory Subcommittee, and has testified on privacy and social policy before the U.S. Department of Health and Human Services (HHS) and the U.S. Food and Drug Administration (FDA). Heywood's work has been profiled in the New Yorker, New York Times Magazine, BusinessWeek, 60 Minutes, CBS Evening News, NPR, Science, and Nature. In 2009, he was chosen for WIRED magazine's "Smart List" and Fast Company's "10 Most Creative People in Healthcare." Heywood and his brother Stephen were the subjects of Pulitzer Prize winner Jonathan Wiener's biography, His Brothers Keeper and the Sundance award-winning documentary, "So Much So Fast."

Vik Kheterpal, MD is a Principal at CareEvolution, Inc., a leading provider of secure interoperability solutions. The company markets HIEBus<sup>™</sup>, a healthcare interoperability platform to enable edge applications to share clinical information in a secure, reliable, and incremental manner. Offering core capabilities like community wide master patient index, terminology standardization, episode grouping, and advanced analytics, HIEBus<sup>™</sup> powers statewide and regional exchanges, regional care coordination networks, provider-centric clinical integration initiatives, and multi-center observational data studies. Dr. Kheterpal is very active in the interoperability and health information technology landscape and serves as Technical Director of the South Carolina Health Information Exchange (SCHIEx). Previously, Dr. Kheterpal served as the global general manager and vice president for clinical information systems for GE Healthcare, where he led GE's clinical IT initiatives. Dr. Kheterpal received his doctorate in medicine from the University of Michigan at Ann Arbor, where he also earned a bachelor's degree in biomedical sciences.

**Rebecca Daniels Kush, PhD** is a Founder and the current President and CEO of CDISC. Dr. Kush has over 25 years of experience in the area of clinical research. She has worked for the U.S. National Institutes of Health, academia, a global contract research organization and pharmaceutical companies in the U.S. and Japan. Among numerous publications, Dr. Kush is lead author of the book, eClinical Trials: Planning and Implementation. Dr. Kush has given invited presentations (including keynotes) and tutorials at industry conferences, FDA and other venues in the U.S., Europe, and Japan for over 20 years. She earned a Ph.D. in Physiology and Pharmacology from the University of California (UCSD) School of Medicine in La Jolla, CA and has a B.S. in Chemistry and Biology from the University of New Mexico.

**Marty LaVenture PhD, MDH** is director of the Office of Health Information Technology and e-Health at the Minnesota Department of Health. Dr. LaVenture is currently leading the statewide Minnesota e-Health Initiative and Directs the Center for Health Informatics at the Department. Current projects include models for e-health profiles, assessment of EHR adoption and work as lead author for the revised chapter on public health informatics in upcoming 4th edition of the Shortliffe & Cimino's *Textbook of Biomedical Informatics (BMI)*. Dr. LaVenture has a master's degree in epidemiology and a PhD in Health Informatics from the University of Minnesota. Previously, he served as the assistant State Epidemiologist for Wisconsin Division of Health and he has also worked for a national private medical software corporation. Dr. LaVenture is currently adjunct member of the faculty at the University of Minnesota in Health Informatics. In 2008 he was named as one of the top 100 influential health leaders in Minnesota. Nationally, Dr. LaVenture serves on the editorial board for the *Journal of Biomedical Informatics*. He is a member of the ASTHO ehealth policy committee; Dr. LaVenture has authored or coauthored many articles and scientific publications, he has delivered numerous presentations to state and national audiences and he has received multiple awards for his work and accomplishments. He is an elected fellow of the American college of medical informatics. Mark Leenay, MS, MD is one of our nation's experts on geriatrics and hospice and palliative care and is the Chief Medical Officer and Senior Vice President at OptumHealth Care Solutions. As chief medical officer, he oversees all clinical programs, ranging from wellness to the most complex medical conditions. He leads the clinical performance team, which is accountable for clinical care and quality in the company's wellness, case management and disease management programs, as well as the clinical performance of external provider partners. Mark's focus is providing members the best care from experienced and knowledgeable providers, leading to shorter hospital stays and improved health. He achieves these goals by championing exceptional performance within OptumHealth's Centers of Excellence and other clinical networks. Previously Mark was chief medical officer for the Medicare and Retirement business for United Healthcare, with accountability for clinical programs, medical payment policy and network relationships. Prior to joining UnitedHealth Group in 2006, Mark directed the palliative care division at the University of Minnesota, Fairview. Mark received his M.D. degree from Thomas Jefferson University and completed his residency at Overlook Hospital, an affiliate of Columbia University. He earned his bachelor's degree from LeMoyne College and his master's degree in psychology from the University of Pennsylvania. Mark is board certified in family medicine, geriatrics, and hospice and palliative care. He is a board member of the Long Term Quality Alliance and sits on the quality and research committees of the National Hospice and Palliative Care Association. He is a former director of the board of the American Academy of Hospice and Palliative Medicine.

**Mia A. Levy, MD, PhD** is the Director of Cancer Clinical Informatics for the Vanderbilt-Ingram Cancer Center and an Assistant Professor of Biomedical Informatics and Medicine. Dr. Levy received her undergraduate degree in Bioengineering from The University of Pennsylvania in 1997 and her Medical Doctorate from Rush University in 2003. She then spent 6 years at Stanford University completing post-graduate training in Internal Medicine and Medical Oncology while completing her PhD in Biomedical Informatics. She joined the faculty at Vanderbilt as an Assistant Professor in Biomedical Informatics and Medicine in August 2009. She is a practicing medical oncologist specializing in the treatment of breast cancer. Dr. Levy's research interests include biomedical informatics methods to support the continuum of cancer care and cancer research. Current research projects include informatics methods for 1) image based cancer treatment response assessment using quantitative imaging, 2) clinical decision support for treatment prioritization of molecular subtypes of cancer, 3) protocol based plan management and 4) learning cancer systems.

**David Madigan, PhD** is Professor and Chair of Statistics at Columbia University in New York City. He received a bachelors degree in Mathematical Sciences and a Ph.D. in Statistics, both from Trinity College Dublin. He has previously worked for AT&T Inc., Soliloquy Inc., the University of Washington, Rutgers University, and SkillSoft, Inc. He has over 100 publications in such areas as Bayesian statistics, text mining, Monte Carlo methods, pharmacovigilance and probabilistic graphical models. He is an elected Fellow of the American Statistical Association and of the Institute of Mathematical Statistics. He has just finished a term as Editor-in-Chief of Statistical Science.

**Carol J. McCall, FSA, MAAA** is the Chief Strategy Officer for GNS Healthcare, a Big Data Analytics company whose industrialized knowledge discovery platform extracts cause-effect relationships directly and at scale from observational data. Her personal goal is to leverage these capabilities to redesign the entire notion of 'evidence' and ignite a true learning system in the healthcare system. Prior to joining GNS Healthcare, Carol was Chief Innovation Officer for Tenzing Health, a subsidiary of Vanguard Health Systems, where she merged creative analytic approaches with human-centered design. Building team-based care models whose approach extended into the community, these approaches were shown to materially improve health, dramatically reduce costs and open new opportunities in a community's economic sustainability. At Humana, Carol led their R&D efforts in their Innovation Center where she pioneered using sophisticated analytics to build a diverse portfolio of prediction, knowledge discovery
and simulation models. She also launched Humana's innovations in personalized medicine, led Humana's Health Services Research Center (HSRC), and helped launch Green Ribbon Health, LLC, a Florida-based company with innovations in health support services for seniors, later serving on its Board of Directors. In other roles at Humana, Carol served as their Chief Information Officer and as VP, Pharmacy Management. Outside of Humana, she served as EVP of Managed Care Business Development for Allscripts Healthcare Solutions and as an actuarial consultant for Milliman, where she helped fashion novel risk-sharing arrangements and implement risk adjustment methodologies. In policy and advisory roles, Carol served as an advisor to the HRP Scientific Program Board, and was a member of the HSRC's governing board. She currently sits on the advisory board of Keas, a consumer health company. Carol is a fellow of the Society of Actuaries and a member of the American Academy of Actuaries.

**Farzad Mostashari, MD, ScM** serves as National Coordinator for Health Information Technology within the Office of the National Coordinator for Health Information Technology at the U.S. Department of Health and Human Services. Farzad joined ONC in July 2009. Previously, he served at the New York City Department of Health and Mental Hygiene as Assistant Commissioner for the Primary Care Information Project, where he facilitated the adoption of prevention-oriented health information technology by over 1,500 providers in underserved communities. Dr. Mostashari also led the Centers for Disease Control and Prevention (CDC) funded NYC Center of Excellence in Public Health Informatics and an Agency for Healthcare Research and Quality funded project focused on quality measurement at the point of care. Prior to this he established the Bureau of Epidemiology Services at the NYC Department of Health, charged with providing epidemiologic and statistical expertise and data for decision making to the health department. Dr. Mostashari did his graduate training at the Harvard School of Public Health and Yale Medical School, internal medicine residency at Massachusetts General Hospital, and completed the CDC's Epidemic Intelligence Service. He was one of the lead investigators in the outbreaks of West Nile Virus and anthrax in New York City, and among the first developers of real-time electronic disease surveillance systems nationwide.

J. Marc Overhage, MD, PhD is the Chief Medical Informatics Officer for Siemens Healthcare. Prior to joining Siemens he was the founding Chief Executive Officer of the Indiana Health Information Exchange and was Director of Medical Informatics at the Regenstrief Institute, Inc., and a Sam Regenstrief Professor of Medical Informatics at the Indiana University School of Medicine. He has spent over 25 years developing and implementing scientific and clinical systems and evaluating their value. With his colleagues from the Regenstrief Institute, he created a community wide electronic medical record (called the Indiana Network for Patient Care) containing data from many sources including laboratories, pharmacies and hospitals in central Indiana. The system currently connects a majority of acute care hospitals in Indiana and includes inpatient and outpatient encounter data, laboratory results, immunization data and other selected data for 12 million patients. In order to create a sustainable financial model, he helped create the Indiana Health Information Exchange, a not-for-profit corporation. In addition Dr. Overhage has developed and evaluated clinical decision support including inpatient and outpatient computerized physician order entry and the underlying knowledge bases to support them. He practiced general internal medicine for over 20 years including the ambulatory, inpatient and emergency care settings. Over the last decade, Dr Overhage has played a significant regional and national leadership role in advancing the policy, standards, financing and implementation of health information exchange. He serves on the Health Information Technology Standards Committee as well as serving on the Board of Directors of the National Quality Form and being engaged in a number of national healthcare initiatives.

**Richard Platt, MD, MSc** is a professor and chair of the Department of Population Medicine at Harvard Medical School and the Harvard Pilgrim Health Care Institute. He is principal investigator of the FDA's Mini-Sentinel program, of contracts with FDA's Center for Drugs Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) to conduct post-marketing studies of drugs' and biologics' safety and effectiveness. He chaired the FDA's Drug Safety and Risk Management Advisory Committee, is a member of the Association of American Medical Colleges' Advisory Panel on Research and the Institute of Medicine Roundtable on Value & Science-Driven Health Care. Dr. Platt was co-chair of the Board of Scientific Counselors of the Centers for Disease Control and Prevention's (CDC) Center for Infectious Diseases. Additionally, he has chaired the National Institutes of Health study section, Epidemiology and Disease Control 2, and the CDC Office of Health Care Partnerships steering committee. Dr. Platt is also principal investigator of a CDC Center of Excellence in Public Health Informatics, the Agency for Healthcare Research and Quality (AHRQ) HMO Research Network Center for Education and Research in Therapeutics, the AHRQ HMO Research Network DEcIDE Center, the CDC Eastern Massachusetts Prevention Epicenter, and FDA contracts to conduct post-marketing studies of drugs' and biologics' safety and effectiveness.



ROUNDTABLE ON VALUE & SCIENCE-DRIVEN HEALTH CARE

## Workshop Logistics

IOM Workshop on Digital Data Priorities for Continuous Learning in Health and Health Care

The Roundtable on Value & Science-Driven Health Care is looking forward to your participation on March 23, 2012. If you have any questions regarding workshop logistics, please contact our office at jcsanders@nas.edu or 202-334-3889.

## LOCATION:

The workshop will be held from 8:00AM – 5:00PM on March 23, 2012 in room 100 of the Keck Center of the National Academies in Washington, DC. The building is located at 500 5th Street, NW. Breakfast will be served at 7:30AM. While the agenda for this meeting has not been finalized, these times provide an accurate estimation for travel planning purposes.

## **DIRECTIONS:**

The meeting site is approximately 5 miles from Washington National Airport and approximately 30 miles from Dulles International Airport. Taxis are most easily hailed on E or F Streets.

The Gallery Place/Chinatown Metro station (YELLOW and GREEN lines) is two blocks away, and only a 15-minute ride from Washington National Airport.

- 1. Exit the station by following signs to Seventh and F Streets/Arena.
- 2. Turn LEFT and walk EAST on F Street NW, two blocks past the Verizon Center.
- 3. Turn RIGHT on to Fifth Street NW
- 4. Walk past the fire station parking lot. The next building on your right will be 500 Fifth St. NW

The **Judiciary Square Metro station (RED line)** is located one block away from the meeting site. Exit the station by following signs to the Building Museum (F Street) exit, between Fourth and Fifth Streets NW

- 1. Turn LEFT and walk WEST on F Street NW
- 2. Cross Fifth Street NW and turn LEFT.
- 3. Walk past the fire station parking lot. The next building on your right will be 500 Fifth St. NW