Re-Use of Operational Electronic Health Record Data for Research and Other Purposes

William Hersh, MD
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: http://informaticsprofessor.blogspot.com

References

Anonymous (2005). Defining the personal health record. *Journal of AHIMA*. 76(6): 24-25. http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_027351.hcsp?dDocName=bok1_027351

Anonymous (2010). Electronic Personal Health Information Exchange: Health Care Entities' Reported Disclosure Practices and Effects on Quality of Care. Washington, DC, Government Accountability Office. http://www.gao.gov/new.items/d10361.pdf

Blumenthal, D (2010). Launching HITECH. *New England Journal of Medicine*. 362: 382-385.

Bourgeois, FC, Olson, KL, et al. (2010). Patients treated at multiple acute health care facilities: quantifying information fragmentation. *Archives of Internal Medicine*. 170: 1989-1995.

Butler, D (2013). When Google got flu wrong. *Nature*. 494: 155-156.

Chapman, WW, Christensen, LM, et al. (2004). Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artificial Intelligence in Medicine*. 33: 31-40.

Denny, JC, Ritchie, MD, et al. (2010). PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 26: 1205-1210.

Finnell, JT, Overhage, JM, et al. (2011). All health care is not local: an evaluation of the distribution of emergency department care delivered in Indiana. *AMIA Annual Symposium Proceedings*, Washington, DC. 409-416.

Gerbier, S, Yarovaya, O, et al. (2011). Evaluation of natural language processing from emergency department computerized medical records for intra-hospital syndromic surveillance. *BMC Medical Informatics & Decision Making*. 11: 50. http://www.biomedcentral.com/1472-6947/11/50

Ginsberg, J, Mohebbi, MH, et al. (2009). Detecting influenza epidemics using search engine query data. *Nature*. 457: 1012-1014.

Greene, SM, Reid, RJ, et al. (2012). Implementing the learning health system: from concept to action. *Annals of Internal Medicine*. 157: 207-210.

Henning, KJ (2004). What is syndromic surveillance? *Morbidity and Mortality Weekly Report*. 53(Suppl): 5-11. http://www.cdc.gov/mmwr/preview/mmwrhtml/su5301a3.htm

Hersh, W (2010). The health information technology workforce: estimations of demands and a framework for requirements. *Applied Clinical Informatics*. 1: 197-212.

Hersh, WR, Cimino, JJ, et al. (2013). Recommendations for the use of operational electronic health record data in comparative effectiveness research. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 1: 14.
http://repository.academyhealth.org/egems/vol1/iss1/14/

Hersh, WR, Weiner, MG, et al. (2013). Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical Care*. 51(Suppl 3): S30-S37.

Jha, AK, Joynt, KE, et al. (2012). The long-term effect of Premier pay for performance on patient outcomes. *New England Journal of Medicine*: Epub ahead of print.

Kann, M and Lewitter, F, Eds. (2013). Translational Bioinformatics. San Francisco, CA, Public Library of Science.

Kern, LM, Malhotra, S, et al. (2013). Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study. *Annals of Internal Medicine*. 158: 77-83.

Kho, AN, Pacheco, JA, et al. (2011). Electronic medical records for genetic research: results of the eMERGE Consortium. *Science Translational Medicine*. 3: 79re71.
http://stm.sciencemag.org/content/3/79/79re1.short

Klompas, M, McVetta, J, et al. (2012). Integrating clinical practice and public health surveillance using electronic medical record systems. *American Journal of Preventive Medicine*. 42(6S2): S154-S162.

Kuperman, GJ (2011). Health-information exchange: why are we doing it, and what are we doing? *Journal of the American Medical Informatics Association*. 18: 678-682.

Lee, SJ and Walter, LC (2011). Quality indicators for older adults: preventing unintended harms. *Journal of the American Medical Association*. 306: 1481-1482.

Longworth, DL (2011). Accountable care organizations, the patient-centered medical home, and health care reform: what does it all mean? *Cleveland Clinic Journal of Medicine*. 78: 571-589.

Miller, HD, Yasnoff, WA, et al. (2009). Personal Health Records: The Essential Missing Element in 21st Century Healthcare. Chicago, IL, Healthcare Information and Management Systems Society.

Nightingale, F (1863). Notes on Hospitals. London, England, Longman, Green, Roberts, Longman, and Green.

Parsons, A, McCullough, C, et al. (2012). Validity of electronic health record-derived quality measurement for performance monitoring. *Journal of the American Medical Informatics Association*. 19: 604-609.

Richesson, RL and Andrews, JE, Eds. (2012). Clinical Research Informatics. New York, NY, Springer.

Safran, C, Bloomrosen, M, et al. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. *Journal of the American Medical Informatics Association*. 14: 1-9.

Selby, JV, Krumholz, HM, et al. (2013). Network news: powering clinical research. *Science Translational Medicine*. 5: 182fs113.
http://stm.sciencemag.org/content/5/182/182fs13.full

Serumaga, B, Ross-Degnan, D, et al. (2011). Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. *British Medical Journal*. 342: d108. http://www.bmj.com/content/342/bmj.d108

Shortliffe, EH (2010). Biomedical informatics in the education of physicians. *Journal of the American Medical Association*. 304: 1227-1228.

Stead, WW, Searle, JR, et al. (2011). Biomedical informatics: changing what physicians need to know and how they learn. *Academic Medicine*. 86: 429-434.

Stecker, EC (2013). The Oregon ACO experiment — bold design, challenging execution. *New England Journal of Medicine*. 368: 982-985.

Tannen, RL, Weiner, MG, et al. (2008). 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. *Pharmacoepidemiology and Drug Safety*. 17: 671-685.

Tannen, RL, Weiner, MG, et al. (2009). Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *British Medical Journal*. 338: b81. http://www.bmj.com/cgi/content/full/338/jan27_1/b81

Tannen, RL, Weiner, MG, et al. (2007). A simulation using data from a primary care practice database closely replicated the Women's Health Initiative trial. *Journal of Clinical Epidemiology*. 60: 686-695.

Wang, TY, Dai, D, et al. (2011). The importance of consistent, high-quality acute myocardial infarction and heart failure care results from the American Heart Association's Get with the Guidelines Program. *Journal of the American College of Cardiology*. 58: 637-644.

Re-Use of Operational Electronic Health Record Data for Research and Other Purposes

William Hersh, MD
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: http://informaticsprofessor.blogspot.com

1

Who said this and when?

- "I am fain to sum up with an urgent appeal for adopting … some uniform system of publishing the statistical records of hospitals. There is a growing conviction that in all hospitals, even in those which are best conducted, there is a great and unnecessary waste of life … In attempting to arrive at the truth, I have applied everywhere for information, but in scarcely an instance have I been able to obtain hospital records fit for any purposes of comparison … If wisely used, these improved statistics would tell us more of the relative value of particular operations and modes of treatment than we have means of ascertaining at present."

2

Who and when...

- Florence Nightingale, *Notes on Hospitals*, London: Longman, Green, Roberts, Longman, and Green, 1863
  - The first informatician?
- She does have a Web site (and museum in London)
  - www.florence-nightingale.co.uk

OREGON HEALTH &SCIENCE UNIVERSITY

Some more recent wisdom

- Stead (2011)
  - Quantity and complexity of information in medicine requires a fundamental paradigm shift as the number of facts per decision rises
- Shortliffe (2010)
  - Focus of medical practice is as much on information as patients, yet we teach very little about it, including its acquisition (EHRs, searching) and use (quality, safety)
- Blumenthal (2010)
  - Information is "the lifeblood of medicine" and health information technology is destined to be "the circulatory system for that information"

OREGON HEALTH &SCIENCE UNIVERSITY

21ˢᵗ century physicians will interact with clinical data in many ways

- In addition to documentation using the EHR, uses include (Safran, 2007)
  - Health information exchange
  - Personal health records
  - Using data to improve care delivery and coordination
  - Quality measurement and improvement
  - Clinical and translational research
  - Public health surveillance
  - Implementing the learning health system

5

OREGON HEALTH &SCIENCE UNIVERSITY

Health information exchange (HIE)

- Patients are "mobile" in many ways – data bears this out
  - In Massachusetts, of 3.69M patients visiting acute care facilities, 31% visited more than one, accounting for 56% of all visits, and 1% visited five or more (Bourgeois, 2010)
  - In Indiana, 40% of patients visiting EDs had data at more than one hospital, with network analysis showed all EDs sharing patients (Finnell, 2011)
- "Data following the patient"
  - Dr. Carolyn Clancy, Director, AHRQ, 2007
- "Electronic sharing of data among hospitals, physicians, clinical laboratories, radiology centers, pharmacies, health plans (insurers), and public health departments." (GAO, 2010)
- Requires that information be interoperable and flow seamlessly across business boundaries (Kuperman, 2011)
- Part of HITECH investment: $564 for state-based HIE

6

OREGON HEALTH &SCIENCE UNIVERSITY

Personal health record (PHR)

- "Electronic lifelong resource of health information needed by individuals to make health decisions," guided by principles (AHIMA, 2005)
  - Individuals own and manage information, which comes from healthcare providers and individual
  - Maintained in secure, private environment
  - Individual determines rights of access
  - Does not replace legal record of provider
- Types of PHRs (Miller, 2009)
  - Tethered – connected to one EHR, e.g., MyChart
  - Standalone – data entered by patient
  - Integrated – data comes from many sources

7

OREGON
HEALTH
&SCIENCE
UNIVERSITY

Using data to improve care delivery and coordination

- US healthcare system still mostly based on fee for service model – little incentive for managing care in coordinated manner
- Primary care medical homes (PCMHs) coordinate care and provide incentive for better use of data (Longworth, 2011)
- Affordable Care Act (ACA, aka Obamacare) implements accountable care organizations (ACOs), which provide bundled payments for conditions (Longworth, 2011)
  - Oregon at forefront with coordinated care organizations (CCOs) (Stecker, 2013)
- All of these innovations require use of data to improve quality and reduce cost

8

OREGON
HEALTH
&SCIENCE
UNIVERSITY

Quality measurement and improvement

- Quality measures increasingly used in US and elsewhere
- Use has been more for process than outcome measures (Lee, 2011), e.g., Stage 1 meaningful use

NQF Measure Number & PQRI Implementation Number	Clinical Quality Measure Title
Core Clinical Quality Measures	
NQF 0013	Hypertension: Blood Pressure Measurement
NQF 0028	Preventive Care and Screening Measure Pair: a) Tobacco Use Assessment, b) Tobacco Cessation Intervention
NQF 0421 PQRI 128	Adult Weight Screening and Follow-up
Alternate Clinical Quality Measures	
NQF 0024	Weight Assessment and Counseling for Children and Adolescents
NQF0041 PQRI 110	Preventive Care and Screening: Influenza Immunization for Patients 50 Years Old or Older
NQF 0038	Childhood Immunization Status

OREGON HEALTH &SCIENCE UNIVERSITY

9

Quality measurement and improvement

- In UK, pay for performance schemes achieved early value but fewer further gains (Serumaga, 2011)
- In US, some quality measures found to lead to improved patient outcomes (e.g., Wang, 2011), others not (e.g., Jha, 2012)
- Desire is to derive automatically from EHR data, but this has proven challenging with current systems (Parsons, 2012; Kern, 2013)

OREGON HEALTH &SCIENCE UNIVERSITY

10

EHR data use for clinical research

- Not only benefit conventional research but allows new approaches (Richesson, 2012), e.g.,
  - Replication of randomized controlled trial (RCT) outcomes using EHR data and statistical corrections (Tannen, 2007; Tannen, 2008; Tannen, 2009)
  - Associating "phenotype" with genotype to replicate known associations as well as identify new ones in eMERGE (Kho, 2011; Denny, 2010)
  - Promise of genomics and bioinformatics yielding other successes as well (Kann, 2013)

11

Public health

- Improving interface between healthcare and public health systems (Klompas, 2012)
- "Syndromic surveillance" – uses data sources for early detection of public health threats, from bioterrorism to emergent diseases
  - Interest increased after 9/11 attacks (Henning, 2004; Chapman, 2004; Gerbier, 2011)
  - One notable success is Google Flu Trends – http://www.google.org/flutrends/
    - search terms entered into Google predict flu activity, but not enough to allow intervention (Ginsberg, 2009)
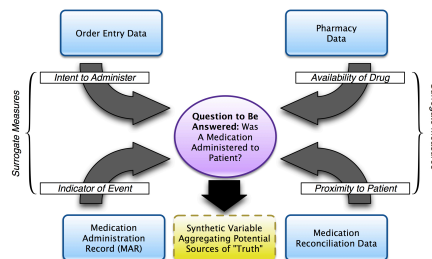   - Less accuracy more recently (Butler, 2013)

12

Implementing the learning healthcare system (Greene, 2012)
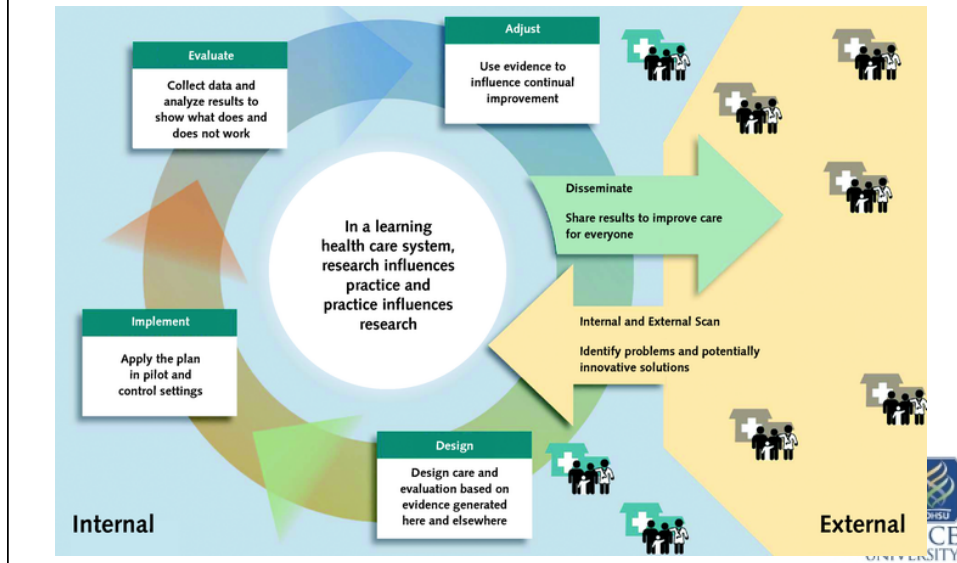
Caveats for use of operational EHR data (Hersh, 2013) – may be

- Inaccurate
- Incomplete
- Transformed in ways that undermine meaning
- Unrecoverable for research
- Of unknown provenance
- Of insufficient granularity
- Incompatible with research protocols

Many "idiosyncrasies" in clinical data (Hersh, 2013)

- "Left censoring" – First instance of disease in record may not be when first manifested
- "Right censoring" – Data source may not cover long enough time interval
- Data might not be captured from other clinical (other hospitals or health systems) or non-clinical (OTC drugs) settings
- Bias in testing or treatment
- Institutional or personal variation in practice or documentation styles
- Inconsistent use of coding or standards

Recommendations for use of operational EHR data (Hersh, 2013)

Recommendation	Description
Apply an Evidence-Based Approach	Ask an answerable question, find the best EHR data ("evidence"), appraise the data, apply evidence to question
Evaluate and Manage Data	Assess availability, completeness, quality (validity), and transformability of data
Create Tools for Data Management	Create software (especially pipelines) for data aggregation, validation and transformation
Determine Metrics for Data Assessment	Determine whether a particular site's data are "research grade"
Develop Methods for Comparative Validation	Develop tools that support analysis of multi-site data collections
Develop a Methodology Knowledge Base	Develop a data catalogue that relates data elements to recommended transformations
Standardize Reporting Methods	Provide details of data sources, provenance and manipulation, to support comparison of data
Engage Informatics Expertise	Ensure validity of findings derived from data collected from disparate sources
Include an Informatics Research Agenda	Generate systematic studies of inherent biases in EHR and data collection methods, such as data entry user interfaces

16

Adapt rules of evidence-based medicine (EBM)?

- Ask an answerable question
  - Can question be answered by the data we have?
- Find the best evidence
  - In this case, the best evidence is the EHR data needed to answer the question
- Critically appraise the evidence
  - Does the data answer our question? Are there confounders?
- Apply it to the patient situation
  - Can the evidence be applied to this setting?

17

OREGON
HEALTH
&SCIENCE
UNIVERSITY

PCORI vision for clinical data research networks (Selby, 2013)

- Adhere to standards and interoperability along the lines of requirements for Stage 2 of meaningful use (e.g., Consolidated CDA, SNOMED, ICD-10, RxNorm, HL7 2.5.1 for public health, etc.)
- Be able to identify cohorts of patients with specific conditions
- Engage patients and clinicians from the health systems in the research process and prioritization
- Develop centralized process for human subjects protection and IRB
- Have a process for contacting patients for health status surveys and recruitment into clinical studies and trials
- Develop a process for outside researchers to access CDRN patients and data
- Embed research activity within functioning healthcare systems without disrupting the business of providing healthcare
- Develop clear and robust policies for privacy and security of data
- Have ability to collect, store, retrieve, process and/or ship biological specimens for research
- Develop an interactive governance process for all of the above

18

OREGON
HEALTH
&SCIENCE
UNIVERSITY

Also need academic contributions of informatics

- Informatics workforce and its training (Hersh, 2010)
  - Development and implementation driven with users and optimal uses in mind – engage by providing value
  - Led by well-trained workforce, including clinical informatics subspecialists
- Research agenda – must understand better (Hersh, 2013)
  - How EHR works and biases health care process creates in its data
  - Workflows – impact and optimization
  - User interfaces that allow the entry of high-quality data in time-efficient manners
  - Limitations of all data and how it can be improved
  - Better adherence to data standards and interoperability

19

Conclusions

- Improving our healthcare system requires information-driven solutions
- There are plentiful opportunities for secondary use or re-use of clinical data
- We must be cognizant of caveats of using operational clinical data
- We must implement best practices for using such data

20