Secondary Use of Clinical Data from Electronic Health Records

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

References


Bernstam, E., Hersh, W., et al. (2009). Synergies and distinctions between computational disciplines in biomedical research: perspective from the Clinical and Translational Science Award programs. Academic Medicine, 84: 964-970.


Secondary Use of Clinical Data from Electronic Health Records

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

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.”
Who and when...

- The first advocate for secondary use of clinical data?
- She does have a Web site (and museum in London)
  - www.florence-nightingale.co.uk

Overview

- Motivations for secondary use of clinical data
- Challenges for secondary use of clinical data
- Types of secondary use of clinical data
  - Clinical and translational research
  - Public health
  - Quality measurement and improvement
Motivations for secondary use of clinical data

- Many “secondary uses” or re-uses of electronic health record (EHR) data, including (Safran, 2007)
  - Clinical and translational research – generating hypotheses and facilitating research
  - Public health surveillance for emerging threats
  - Healthcare quality measurement and improvement
- Opportunities facilitated by growing incentives for “meaningful use” of EHRs in the HITECH Act (Blumenthal, 2011; Blumenthal, 2011), aiming toward the “learning healthcare system” (Eden, 2008; Friedman, 2010)

Challenges for secondary use of clinical data

- EHR data does not automatically lead to knowledge
  - Data quality and accuracy is not a top priority for busy clinicians (de Lusignan, 2005)
  - There is “tension” between structured and narrative documentation (Rosenbloom, 2011)
- Many data idiosyncrasies (Weiner, 2011)
  - “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
Data in EHRs is incomplete

- Claims data failed to identify more than half of patients with prognostically important cardiac conditions prior to admission for catheterization (Jollis, 1993)
- Various approaches generated variable rate of retrieval of cases for quality measurement (Benin, 2005; Rhodes, 2007; Parsons, 2012); algorithmic methods can lead to improvement (Benin, 2011)
- At Columbia University Medical Center, 48.9% of patients with ICD-9 code for pancreatic cancers did not have corresponding disease documentation in pathology reports, with many data elements incompletely documented (Botsis, 2010)

Patients get care in multiple places

- Study of 3.7M patients in Massachusetts found 31% visited 2 or more hospitals over 5 years (57% of all visits) and 1% visited 5 or more hospitals (10% of all visits) (Bourgeois, 2010)
- Study of 2.8M emergency department (ED) patients in Indiana found 40% of patients had data at multiple institutions, with all 81 EDs sharing patients in common (Finnell, 2011)
Despite these caveats, can we use secondary data from EHRs?

- Clinical and translational research
- Public health
- Quality measurement and improvement

Clinical research

- Clinical research comprises studies and trials in human subjects that fall into the three sub-categories
  - Patient-oriented research – research conducted with human subjects (or on material of human origin such as tissues, specimens and cognitive phenomena) for which an investigator (or colleague) directly interacts with human subjects
  - Epidemiologic and behavioral studies
  - Outcomes research and health services research
Translational research (Zerhouni, 2007)

- Accelerating research results from laboratory to clinical environment to community along T1/T2/T3 axis (Dougherty, 2008)

Clinical research informatics (CRI)

- CRI is the “application of informatics principles and techniques to support the spectrum of activities and business processes that instantiate clinical research.” (Richesson, 2012)
- Historical focus on management of research protocols and data capture and analysis, but increasingly focuses on integration of clinical systems and secondary use of clinical data (Embi, 2009; Payne, 2009)
- Another area where there is distinction between informatics and related disciplines (e.g., computer science) (Bernstam, 2009)
Informatics can help achieve integration of:
• Research systems
• Research activities
• Clinical systems (Payne, 2005)

Clinical & Translational Science Award (CTSA) Program

• www.ctsacentral.org
• Goal is to accelerate translation of research into clinical care and community
• Funding 60 centers around country in pursuit of goal to improve human health by
  – Streamlining science
  – Transforming training environments
  – Improving the conduct, quality and dissemination of clinical and translational research
• Is informatics important?
  – Centers required to have a biomedical informatics component
  – The word “informatics” appeared 34 times in original Request for Applications (RFA)!
Common themes emerging among CTSA informatics programs

- Web portal to integrate research activities, people, etc.
- Informatics services and consultation for investigators
- Data warehousing for clinical, laboratory, and/or specimen data
  - Long-term but difficult goal: data sharing
  - Also difficult but being undertaken by many: clinical data warehouses from EHRs and other systems
- Facilitating collaboration and translation
- Educational programs for translational researchers
  - Some programs also focused on education for informaticians

Contents of research data warehouses in CTSA programs (MacKenzie, 2012)
Examples of research results from clinical data

- Identifying patients with heart failure for clinical research (Pakhomov, 2007)
  - Used natural language processing and predictive modeling to identify patients with heart failure for research studies
- Relationship of childhood obesity to hyperglycemia in pregnancy (Hillier, 2007)
  - Using data for nearly 10,000 patients, found association between hyperglycemia in pregnancy and subsequent childhood obesity
- “Replication” of clinical trials results by using techniques and measures, e.g., Women’s Health Initiative (Tannen, 2007) and hypertension studies (Tannen, 2008)

Including use of genomic data

- Electronic Medical Records and Genomics (eMERGE) Network – consortium that aims to link data in biorepositories with EHR systems (including clinical text processed by NLP) for “large-scale, high-throughput genetic research” (McCarty, 2011; Wilke, 2011)
- Results include
  - Replicating finding of known gene-disease associations from research data in EHR data for several diseases (Denny, 2010; Ritchie, 2010)
  - Discovering new gene-disease associations (Denny, 2010)
  - Detection of clinical events, e.g., detection of electrocardiogram QT interval prolongation (Denny, 2009) and associated genomic predictors (Denny, 2010)
Public health

- Public health performs its missions through its core functions
  - Assessment
  - Policy Development
  - Assurance
- Public health activities include
  - Prevent epidemics and the spread of disease
  - Protect against environmental hazards
  - Prevent injuries
  - Promote and encourage healthy behaviors
  - Respond to disasters and assists communities in recovery
  - Assure the quality and accessibility of health services
- IOM (2010)
- www.whatispublichealth.org

Early application of public health informatics

- http://www.ph.ucla.edu/epi/snow.html
- John Snow was an early epidemiologist in the UK in the mid-19th century
- In 1854, he investigated a rapid outbreak of cholera in Soho area of London
- He found a common characteristic of those infected: use of water from the Broad Street pump
- Dots represented cases; X’s represented clusters
- A sort of early use of a geographic information system (GIS)
Despite laws, there is inadequate completeness of reporting

- Systematic review of 33 published reports between 1970-1999 found variation from 9% to 99%, with completeness for AIDS, sexually transmitted infections, and TB higher (79%) than for all other diseases combined (49%) (Doyle, 2002)

- Reasons for clinicians not reporting
  - Unaware of legal requirement
  - Lack of knowledge of which diseases reportable
  - Do not understand how to report
  - Assumption that someone else will report
  - Intentional failure for privacy reasons
Efforts to improve reporting

• Emerging strategies for increasing reporting (Silk, 2005)
  – Active surveillance when appropriate
  – Automated, electronic laboratory-based reporting
  – Strengthening ties with clinicians and other key partners
  – Increasing use of laboratory diagnostic tests in identifying new cases
• Automated electronic laboratory reporting
  – Shown to increase rate of reporting in a variety of settings (e.g., Effler, 1999; Panackal, 2002)
  – Adding data from an electronic health information exchange identified 4.4 times as many cases as spontaneous paper-based methods and identified those cases 7.9 days earlier (Overhage, 2008)

Amount of data to aid surveillance is increasing

• EHR data – e.g.,
  – Radiology reports (Chapman, 2003)
  – Chief complaints (Chapman, 2004)
  – Emergency Department chief complaint text (Chapman, 2005)
  – Laboratory ordering and results, e.g., lumbar puncture (Kimia, 2006)
  – All of the above (Klompas, 2012)
• Purchase of over-the-counter medications (e.g., cough suppressants, antidiarrheal medications, etc.) can foretell epidemics (Proctor, 1998; Goldenberg, 2002; Hogan, 2003)
Even more surveillance data in modern times

- Search engines
  - First major use: 2002 Winter Olympics in Salt Lake City, UT (Johnson, 2004)
  - Google Flu Trends: http://www.google.org/flutrends/
  - Some flu-related queries track closely with flu-related visits to physicians, lagging by about a day (Ginsberg, 2009)
  - Also shown with a search log submitted to a Swedish Web site (Hulth, 2009)
- Social networks, e.g., Facebook (Christakis, 2010)
- Practical value?

Healthcare quality

- The Berwick “triple aim” (Berwick, 2008)
  - Improving experience of care
  - Improving health of population
  - Reducing costs
- Healthcare must strive to be “high-value, cost-conscious” (Owens, 2011), and “high reliability” (Chassin, 2011)
  - An essential physician competency (Weinberger, 2011) and should be part of continuing medical education (Shojania, 2012)
- Investments in health IT can enable the “learning health system” (Friedman, 2010), allowing us to take advantage of data sources that already exist, such as administrative sources, registries, laboratories, and electronic health records (EHRs) (Roski, 2011)
Donabedian (2002) model of quality

- Three categories
  - Structural – factors that make it easier or harder to deliver high-quality care, e.g., hospital location, volume, physician licensure, nurse staffing levels
  - Process – factors describing healthcare content and activities, e.g., adherence to guidelines for screening, treatment, etc.
  - Outcomes – changes attributable to care, e.g., mortality, morbidity, functional status
- Implemented and measured at different levels at an institution, e.g., individual, department, organization

Other definitions and issues

- Pay for performance (P4P) (Maynard, 2012)
  - Based on notion that healthcare should be held accountable financially and otherwise
  - Sometimes called value-based purchasing (Leapfrog, 2007)
- Public reporting (Mehrotra, 2012)
  - Publishing of information about quality, cost, and other aspects of healthcare performance
- Quality improvement collaboratives (O’Connor, 2012)
  - Putting the learning health system into action (Greene, 2012)
EHRs can augment data used in quality measures

- Coded information in EHR
  - Improves ability to assess diabetes quality measures (Tang, 2007)
  - Administrative (or “claims”) data insufficient to calculate HEDIS measures – EHR data can improve accuracy of calculating HEDIS measures (Pawlson, 2007)
- Overall, EHR data quality is mixed for quality measurement; important attributes to improve include are granularity, timeliness, and comparability (Chan, 2010)
- EHR data for measuring quality requires transitions in collection and transformation of data (Weiner, 2012)

EHRs and quality measures

- Some measures in narrative text are harder to access
  - In heart failure, important data inaccessible in clinical notes, especially exclusion data for medications (Baker, 2007)
  - Some data can be extracted by natural language processing (NLP) as effectively as manual abstractors in areas such smoking cessation advice (Hazlehurst, 2005), diabetic foot exam (Pakhomov, 2008), and congestive heart failure (CHF) (Pakhomov, 2008)
- Also impacted by workflow and documentation practices, which can result in underreporting of some preventive measures (Parsons, 2012)
Most prominent quality measures come from HITECH

• Eligible Professional (outpatient) – three required or alternate measures plus three of 13 others, e.g.,
  – Hypertension – blood pressure measurement
  – Tobacco use assessment and cessation intervention
  – Adult weight screening and follow-up
• Eligible Hospital (inpatient) – 15 required measures, e.g.,
  – Diabetes: Hemoglobin A1c, low-density lipoprotein, and blood pressure control
  – Influenza immunization for patients > 50 years old
  – Pneumonia vaccination status for older adults
  – Breast cancer screening
  – Colorectal cancer screening

Secondary use of EHR data

• Offers potential to improve
  – Clinical and translational research
  – Public health
  – Quality measurement and improvement
• HITECH SHARP initiative investing in further work (Chute, 2011; Rea, 2012)
• Also many challenges
  – Quality and completeness of data – how to improve what clinicians enter?
  – Privacy – how to protect?
  – Cost – who pays?