Big Data in Healthcare and Biomedicine: Opportunities and Challenges

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

References
http://healthland.time.com/2013/02/20/bitter-pill-why-medical-bills-are-killing-us/


Classen, DC, Resar, R, et al. (2011). 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Affairs.* 30: 4581-4589.


Big Data in Healthcare and Biomedicine: Opportunities and Challenges

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

Goal and outline

• Goal
  – Provide a balanced view of opportunities and limitations of Big Data in healthcare and biomedicine

• Agenda
  – Growing data in healthcare
  – Opportunities for secondary use or re-use of clinical data for research and other purposes
  – Caveats of using operational clinical data
  – Big Data in Healthcare at OHSU
    • Biomedical Informatics educational program
    • Big Data to Knowledge (BD2K) projects
    • Informatics Discovery Lab (IDL)
Many problems in healthcare have information-related solutions

- Quality – not as good as it could be (McGlynn, 2003; Schoen, 2009; NCQA, 2010)
- Safety – errors cause morbidity and mortality; many preventable (Kohn, 2000; Classen, 2011; van den Bos, 2011; Smith, 2012)
- Cost – rising costs slowing, but US still spends more but gets less (Angrisano, 2007; Brill, 2013; Hartman, 2015)
- Inaccessible information – missing information frequent in primary care (Smith, 2005)

Growing evidence that information interventions are part of solution

- Systematic reviews (Chaudhry, 2006; Goldzweig, 2009; Buntin, 2011; Jones, 2014) have identified benefits in a variety of areas, although
  - Quality of many studies could be better
  - Large number of early studies came from a small number of “health IT leader” institutions

(Buntin, 2011)
US has made substantial investment in health information technology (HIT)

“To improve the quality of our health care while lowering its cost, we will make the immediate investments necessary to ensure that within five years, all of America’s medical records are computerized ... It just won’t save billions of dollars and thousands of jobs – it will save lives by reducing the deadly but preventable medical errors that pervade our health care system.”

January 5, 2009

Health Information Technology for Economic and Clinical Health (HITECH) Act of the American Recovery and Reinvestment Act (ARRA) (Blumenthal, 2011)

- Incentives for electronic health record (EHR) adoption by physicians and hospitals (up to $27B)
- Direct grants administered by federal agencies ($2B, including funding for health information exchange, workforce development, etc.

Which has led to significant EHR adoption in the US

- Office-based physicians (Hsiao, 2014)
- Outpatient departments (Jamoom, 2015)
- Emergency departments (Jamoom, 2015)
- Non-federal hospitals (Charles, 2014)
Providing opportunities for “secondary use” or “re-use” of clinical data

- (Safran, 2007; SHARPn, Rea, 2012)
- Using data to improve care delivery
- Healthcare quality measurement and improvement
- Clinical and translational research
- Public health surveillance
- Implementing the learning health system

Complemented by research data

- NIH Clinical and Translational Science Award (CTSA) Program
  - [www.ctsacentral.org](http://www.ctsacentral.org)
  - Growing numbers of research data warehouses (Mackenzie, 2012) and secondary use approaches (Vanderbilt; Danciu, 2014)
- Growing availability and quantity of “omics” data
  - Genome, metabolome, interactome, etc. (Witten, 2013)
- New opportunities in “precision medicine” (IOM, 2011; Collins, 2015; Kaiser, 2015)
CTSA funding leading to creation of research data warehouses

(Mackenzie, 2012)

Complemented by growing amounts of research data

(Danciu, 2014)
And falling cost of “omics” data

http://www.genome.gov/sequencingcosts/

How are we using this (big) data?

• Improving healthcare
• Advancing biomedical research
• Public health
Using data to improve healthcare

• With shift of payment from “volume to value,” healthcare organizations will need to manage information better to provide better care (Diamond, 2009; Horner, 2012)

• Predictive analytics is use of data to anticipate poor outcomes or increased resource use – applied by many to problem of early hospital re-admission (e.g., Gildersleeve, 2013; Amarasingham, 2013; Herbert, 2014)

• Also can be used to measure quality of care delivered to make it more “accountable” (Hussey, 2013; Barkhuysen, 2014)
Clinical and translational research (cont.)

• One of largest and most productive efforts has been eMERGE Network – connecting genotype-phenotype (Gottesman, 2013; Newton, 2013)
  – [http://emerge.mc.vanderbilt.edu](http://emerge.mc.vanderbilt.edu)
  – Has used EHR data to identify genomic variants associated with various phenotypes (Denny, 2012; Denny, 2013)

• Much potential for using observational studies as complement to randomized controlled trials (RCTs) (Dahabreh, 2014)

Public health

• “Syndromic surveillance” aims to use 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)

• Ongoing effort in Google Flu Trends
  – [http://www.google.org/flutrends/](http://www.google.org/flutrends/)
  – Search terms entered into Google predicted flu activity but not early enough to intervene (Ginsberg, 2009)
  – Performance in recent years has been poorer (Butler, 2013)
  – Case of needing to avoid “Big Data hubris” (Lazer, 2014)
Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research

William R. Hersh, MD,* Mark G. Weiner, MD,† Peter J. Embi, MD, MS; Judith R. Logan, MD, MS; Philip R. Payne, PhD,‡ Elmer V. Bernstein, MD, MSE; Harold P. Lehmann, MD, PhD; George Hripcsak, MD, MS; Timothy H. Hartzog, MD, MS; James J. Cimino, MD; and Joel H. Saltz, MD, PhD

Operational clinical data may be (Medical Care, 2013):

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

Caveats of clinical data

• Documentation not always a top priority for busy clinicians (de Lusignan, 2005)
• Not every diagnosis is recorded at every visit; absence of evidence is not always evidence of absence, an example of a concern known by statisticians as censoring (Zhang, 2010)
• Makes seemingly simple tasks such as identifying diabetic patients challenging (Miller, 2004; Wei, 2013; Richesson, 2013)
"Idiosyncrasies" of 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

A big challenge is interoperable data
Challenges to EHRs and HIE have spurred focus on interoperability

- Office of National Coordinator for Health IT (ONC) developing interoperability road map for 10-year path forward (Galvez, 2014)
- Emerging approaches include
  - RESTful architectures for efficient client-server interaction
  - OAuth2 for Internet-based security
  - Standard application programming interface (API) for query/retrieval of data
    - Need for both documents and discrete data
    - Emerging standard is Fast Health Interoperability Resources (FHIR)

Also need to develop clinical data research networks

- Established
  - HMO Research Network – facilitates clinical research
    - [www.hmoresearchnetwork.org](http://www.hmoresearchnetwork.org)
  - FDA Mini-Sentinel Network – safety surveillance
    - [www.mini-sentinel.org](http://www.mini-sentinel.org)
- New
  - PCORnet – [www.pcornet.org](http://www.pcornet.org)
    - Clinical data research networks (CDRNs) – 11 networks aggregating data on >1M patients each
      - (Fleurence, 2014; Collins, 2014; and other papers in JAMIA special issue)
    - Common Data Model for subset of data
Large amounts of data do not obviate the need for experimental research


- In large analysis of clinical trials, about slightly more than half of new treatments were superior to established ones (Djulbegovic, 2012)
- GWAS studies have yet to lead to substantial clinical insights (Visscher, 2012)

A final concern: privacy

- Re-identifying data is a long-known problem
  - Famous case of identifying Governor of Massachusetts from public data sources (Sweeney, 2002)
- Data breaches are worsening with proliferation of EHR systems
  - Theft of 80 million records from Anthem insurer earlier in 2015 (Abelson, 2015)
- Growing tracking of our data makes the task even easier
  - Re-identification via credit card transactions (de Montjoye, 2015)
Big Data opportunities at OHSU

• Biomedical Informatics educational program
• Big Data to Knowledge (BD2K) projects
• Informatics Discovery Lab (IDL)

Biomedical Informatics Graduate Program

• Opportunities for future professionals, leaders, and researchers (Hersh, 2010)
• Graduate-level programs
  – Graduate Certificate
  – Master’s – research, professional
  – PhD
• Graduate Certificate and Master’s available online
• Innovations in online learning, including AMIA 10x10 Program

http://www.ohsu.edu/informatics
Big Data to Knowledge (BD2K) projects

• NIH program
  – http://bd2k.nih.gov

• Programs funded
  – Centers of Excellence
  – Education and Training
    • Two OHSU awards
      – Development of Open Educational Resources
      – Big Data Skills Course

• Non-BD2K data-related projects: OHSU collaboration with Mayo Clinic in large-scale processing of EHR data for cohort discovery

Informatics Discovery Lab: academia-industry collaboration

http://www.ohsu.edu/idl
Conclusions

- 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
- We need consensus on approaches to standards and interoperability
- There are opportunities for rewarding careers for diverse professionals

For more information

- Bill Hersh
  - http://www.billhersh.info
- Informatics Professor blog
  - http://informaticsprofessor.blogspot.com
- OHSU Department of Medical Informatics & Clinical Epidemiology (DMICE)
  - http://www.ohsu.edu/informatics
  - http://www.youtube.com/watch?v=T-74duDDwU
  - http://oninformatics.com
- What is Biomedical and Health Informatics?
  - http://www.billhersh.info/whatis
- Office of the National Coordinator for Health IT (ONC)
  - http://healthit.hhs.gov
- American Medical Informatics Association (AMIA)
  - http://www.amia.org
- National Library of Medicine (NLM)