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

Abelson, R and Creswell, J (2015). Data Breach at Anthem May Forecast a Trend. New York, NY. New York Times. February 6, 2015. http://www.nytimes.com/2015/02/07/business/data-breach-at-anthem-may-lead-to-others.html

Amarasingham, R, Patel, PC, et al. (2013). Allocating scarce resources in real-time to reduce heart failure readmissions: a prospective, controlled study. *BMJ Quality & Safety*. 22: 998-1005.

Angrisano, C, Farrell, D, et al. (2007). Accounting for the Cost of Health Care in the United States. Washington, DC, McKinsey & Company. http://www.mckinsey.com/mgi/rp/healthcare/accounting_cost_healthcare.asp

Anonymous (2010). The State of Health Care Quality: 2010. Washington, DC, National Committee for Quality Assurance. http://www.ncqa.org/tabid/836/Default.aspx

Anonymous (2011). Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, DC, National Academies Press.

Barkhuysen, P, deGrauw, W, et al. (2014). Is the quality of data in an electronic medical record sufficient for assessing the quality of primary care? *Journal of the American Medical Informatics Association*. 21: 692-698.

Blumenthal, D (2011). Implementation of the federal health information technology initiative. *New England Journal of Medicine*. 365: 2426-2431.

Blumenthal, D (2011). Wiring the health system--origins and provisions of a new federal program. *New England Journal of Medicine*. 365: 2323-2329.

Brill, S (2013). Bitter Pill: Why Medical Bills Are Killing Us. Time, April 4, 2013. http://healthland.time.com/2013/02/20/bitter-pill-why-medical-bills-are-killing-us/

Buntin, MB, Burke, MF, et al. (2011). The benefits of health information technology: a review of the recent literature shows predominantly positive results. *Health Affairs*. 30: 464-471.

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.
Charles, D, Gabriel, M, et al. (2014). Adoption of Electronic Health Record Systems among U.S. Non-federal Acute Care Hospitals: 2008-2013. Washington, DC, Department of Health and Human Services. http://www.healthit.gov/sites/default/files/oncdatabrief16.pdf
Chaudhry, B, Wang, J, et al. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine*. 144: 742-752.
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.
Collins, FS, Hudson, KL, et al. (2014). PCORnet: turning a dream into reality. *Journal of the American Medical Informatics Association*. 21: 576-577.
Collins, FS and Varmus, H (2015). A new initiative on precision medicine. *New England Journal of Medicine*: Epub ahead of print.
Dahabreh, IJ and Kent, DM (2014). Can the learning health care system be educated with observational data? *Journal of the American Medical Association*. 312: 129-130.
Danciu, I, Cowan, JD, et al. (2014). Secondary use of clinical data: the Vanderbilt approach. *Journal of Biomedical Informatics*: Epub ahead of print.
deLusignan, S and vanWeel, C (2005). The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice*. 23: 253-263.
deMontjoye, YA, Hidalgo, CA, et al. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*. 3: 1376.
http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html
Denny, JC (2012). Mining Electronic Health Records in the Genomics Era. PLOS Computational Biology: Translational Bioinformatics. M. Kann and F. Lewitter.
Denny, JC, Bastarache, L, et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature Biotechnology*. 31: 1102-1111.
Diamond, CC, Mostashari, F, et al. (2009). Collecting and sharing data for population health: a new paradigm. *Health Affairs*. 28: 454-466.
Djulbegovic, B, Kumar, A, et al. (2012). New treatments compared to established treatments in randomized trials. *Cochrane Database of Systematic Reviews*. 10: MR000024.
Fleurence, RL, Curtis, LH, et al. (2014). Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association*. 21: 578-582.
Galvez, E (2014). Interoperability Roadmap: Early Draft Material - Joint HITPC and HITSC Meeting. Washington, DC, Department of Health and Human Services.
http://www.healthit.gov/facas/sites/faca/files/Joint_HIT_DRAFT Nationwide Interoperability Roadmap Material FACA v2_2014-10-15.pdf
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

Gildersleeve, R and Cooper, P (2013). Development of an automated, real time surveillance tool for predicting readmissions at a community hospital. *Applied Clinical Informatics*. 4: 153-169.
Ginsberg, J, Mohebbi, MH, et al. (2009). Detecting influenza epidemics using search engine query data. *Nature*. 457: 1012-1014.
Goldzweig, CL, Towfigh, A, et al. (2009). Costs and benefits of health information technology: new trends from the literature. *Health Affairs*. 28: w282-w293.
Gottesman, O, Kuivaniemi, H, et al. (2013). The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genetics in Medicine*. 15: 761-771.
Hebert, C, Shivade, C, et al. (2014). Diagnosis-specific readmission risk prediction using electronic health data: a retrospective cohort study. *BMC Medical Informatics & Decision Making*. 14: 65. http://www.biomedcentral.com/1472-6947/14/65
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, 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.
Horner, P and Basu, A (2012). Analytics & the future of healthcare. Analytics, January/February 2012. http://www.analytics-magazine.org/januaryfebruary-2012/503-analytics-a-the-future-of-healthcare
Hsiao, CJ and Hing, E (2014). Use and Characteristics of Electronic Health Record Systems Among Office-based Physician Practices: United States, 2001–2013. Hyattsville, MD, National Center for Health Statistics, Centers for Disease Control and Prevention. http://www.cdc.gov/nchs/data/databriefs/db143.htm
Hussey, PS, Wertheimer, S, et al. (2013). The association between health care quality and cost: a systematic review. *Annals of Internal Medicine*. 158: 27-34.
Jamoom, E and Hing, E (2015). Progress With Electronic Health Record Adoption Among Emergency and Outpatient Departments: United States, 2006–2011. Hyattsville, MD National Center for Health Statistics, Centers for Disease Control and Prevention. http://www.cdc.gov/nchs/data/databriefs/db187.htm
Jones, SS, Rudin, RS, et al. (2014). Health information technology: an updated systematic review with a focus on meaningful use. *Annals of Internal Medicine*. 160: 48-54.
Kaiser, J (2015). NIH plots million-person megastudy. *Science*. 347: 817.
Kohn, LT, Corrigan, JM, et al., Eds. (2000). To Err Is Human: Building a Safer Health System. Washington, DC, National Academies Press.
Lazer, D, Kennedy, R, et al. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science*. 343: 1203-1205.
MacKenzie, SL, Wyatt, MC, et al. (2012). Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. *Journal of the American Medical Informatics Association*. 19(e1): e119-e124.
McGlynn, EA, Asch, SM, et al. (2003). The quality of health care delivered to adults in the United States. *New England Journal of Medicine*. 348: 2635-2645.
Miller, DR, Safford, MM, et al. (2004). Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care*. 27(Suppl 2): B10-21.

Newton, KM, Peissig, PL, et al. (2013). Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *Journal of the American Medical Informatics Association*. 20(e1): e147-154.
Prasad, V, Vandross, A, et al. (2013). A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clinic Proceedings*. 88: 790-798.
Rea, S, Pathak, J, et al. (2012). Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project. *Journal of Biomedical Informatics*. 45: 763-771.
Richesson, RL, Rusincovitch, SA, et al. (2013). A comparison of phenotype definitions for diabetes mellitus. *Journal of the American Medical Informatics Association*. 20(e2): e319-e326.
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.
Schoen, C, Osborn, R, et al. (2009). A survey of primary care physicians in eleven countries, 2009: perspectives on care, costs, and experiences. *Health Affairs*. 28: w1171-1183.
Smith, M, Saunders, R, et al. (2012). Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, DC, National Academies Press.
Smith, PC, Araya-Guerra, R, et al. (2005). Missing clinical information during primary care visits. *Journal of the American Medical Association*. 293: 565-571.
Sweeney, L (2002). k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*. 10: 557-570.
VanDenBos, J, Rustagi, K, et al. (2011). The $17.1 billion problem: the annual cost of measurable medical errors. *Health Affairs*. 30: 596-603.
Visscher, PM, Brown, MA, et al. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*. 90: 7-24.
Wei, WQ, Leibson, CL, et al. (2013). The absence of longitudinal data limits the accuracy of high-throughput clinical phenotyping for identifying type 2 diabetes mellitus subjects. *International Journal of Medical Informatics*. 82: 239-247.
Witten, DM and Tibshirani, R (2013). Scientific research in the age of omics: the good, the bad, and the sloppy. *Journal of the American Medical Informatics Association*. 20: 125-127.
Zhang, Z and Sun, J (2010). Interval censoring. *Statistical Methods in Medical Research*. 19: 53-70.

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)

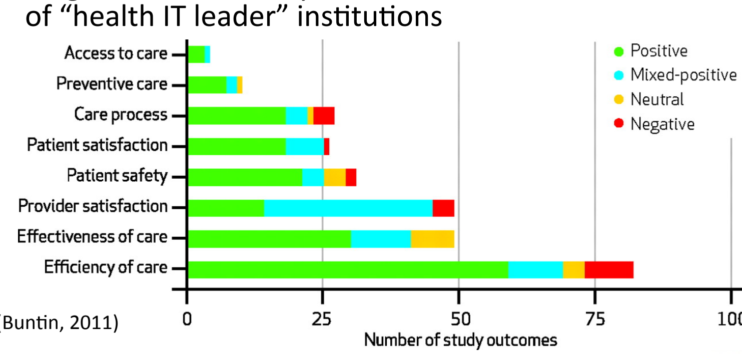
Many problems in healthcare have information-related solutions

- 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)

OREGON HEALTH & SCIENCE UNIVERSITY

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)

OREGON HEALTH & SCIENCE UNIVERSITY

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)

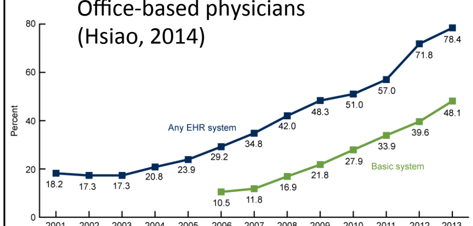
- Incentives for electronic health record (EHR) adoption by physicians and hospitals (up to $27B)
- Direct grants administered by federal agencies ($2B, including funding for health information exchange, workforce development, etc.

OREGON HEALTH & SCIENCE UNIVERSITY

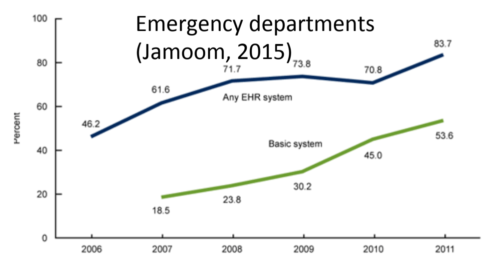
5

Which has led to significant EHR adoption in the US

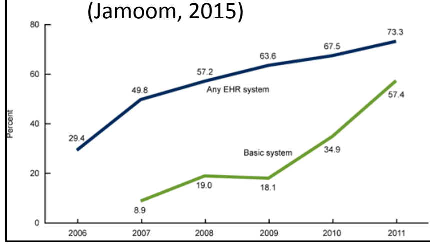
Office-based physicians (Hsiao, 2014)



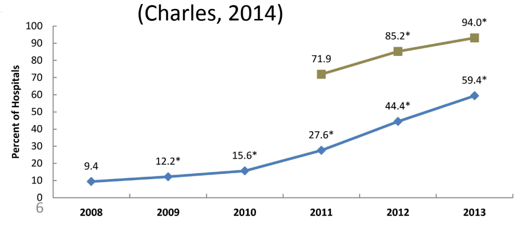
Emergency departments (Jamoom, 2015)



Outpatient departments (Jamoom, 2015)



Non-federal hospitals (Charles, 2014)



6

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

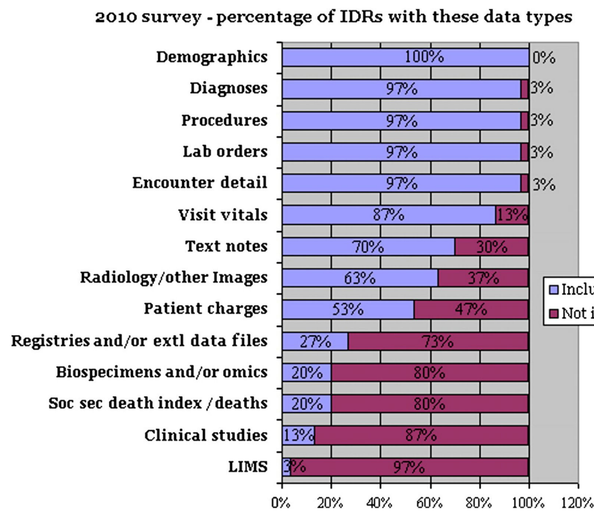
7

Complemented by research data

- NIH Clinical and Translational Science Award (CTSA) Program
  - 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)

8

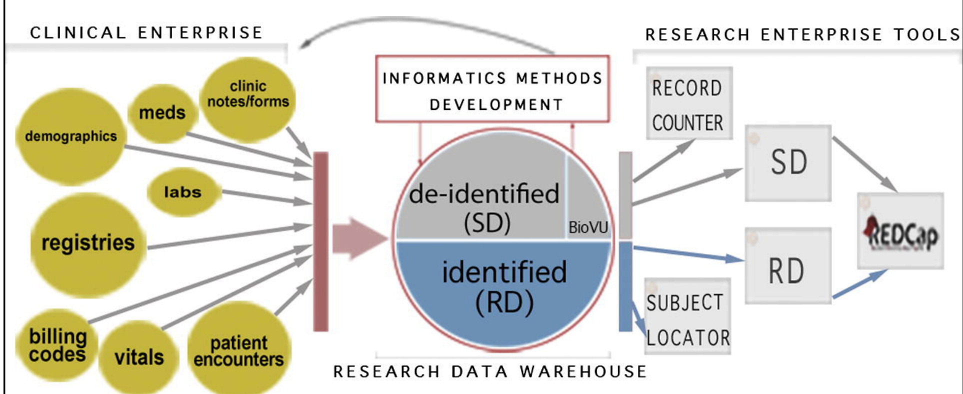
CTSA funding leading to creation of research data warehouses



(Mackenzie, 2012)

9

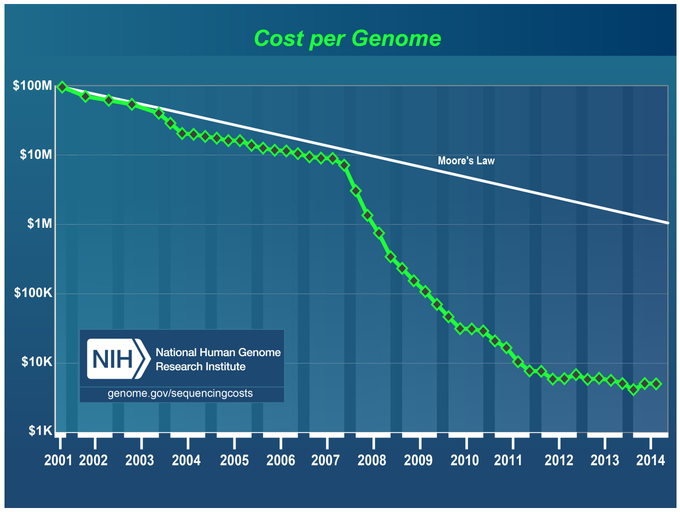
Complemented by growing amounts of research data



(Danciu, 2014)

10

And falling cost of "omics" data



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

OREGON HEALTH & SCIENCE UNIVERSITY

How are we using this (big) data?

- Improving healthcare
- Advancing biomedical research
- Public health

OREGON HEALTH & SCIENCE UNIVERSITY

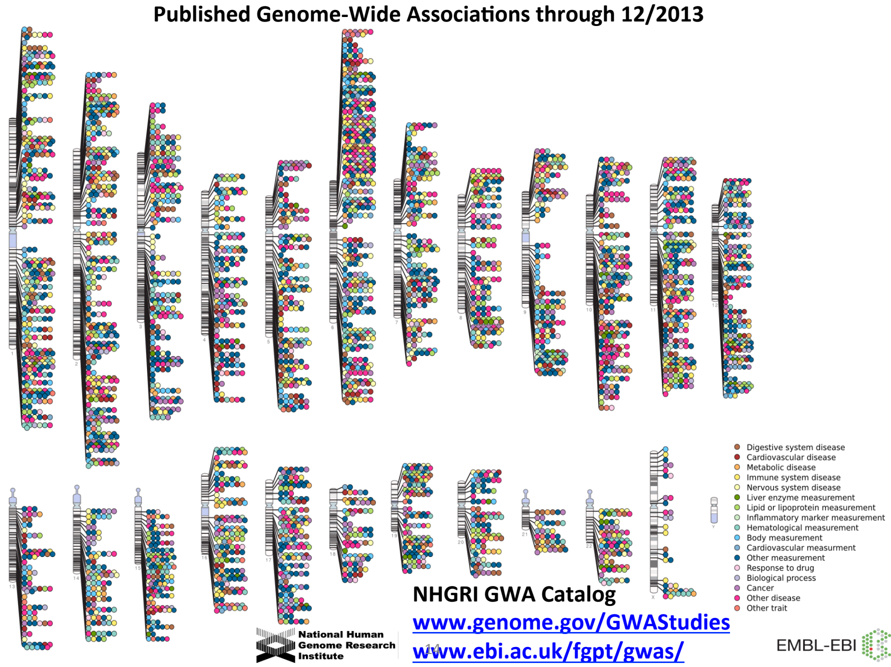
Using data to improve healthcare

- 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)

13

OREGON HEALTH & SCIENCE UNIVERSITY

Advancing Biomedical Research:
Published Genome-Wide Associations through 12/2013



NHGRI GWA Catalog
www.genome.gov/GWAStudies
www.ebi.ac.uk/fgpt/gwas/

National Human Genome Research Institute

EMBL-EBI

14

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
  - 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)

OREGON HEALTH & SCIENCE UNIVERSITY

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/
  - 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)

OREGON HEALTH & SCIENCE UNIVERSITY

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.O. Payne, PhD,‡ Elmer V. Bernstam, 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

(*Med Care* 2013;00: 000–000)

Health Information Technology (ONC) through the Strategic Health IT Advanced Research Projects (SHARP) Program,

17

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)

18

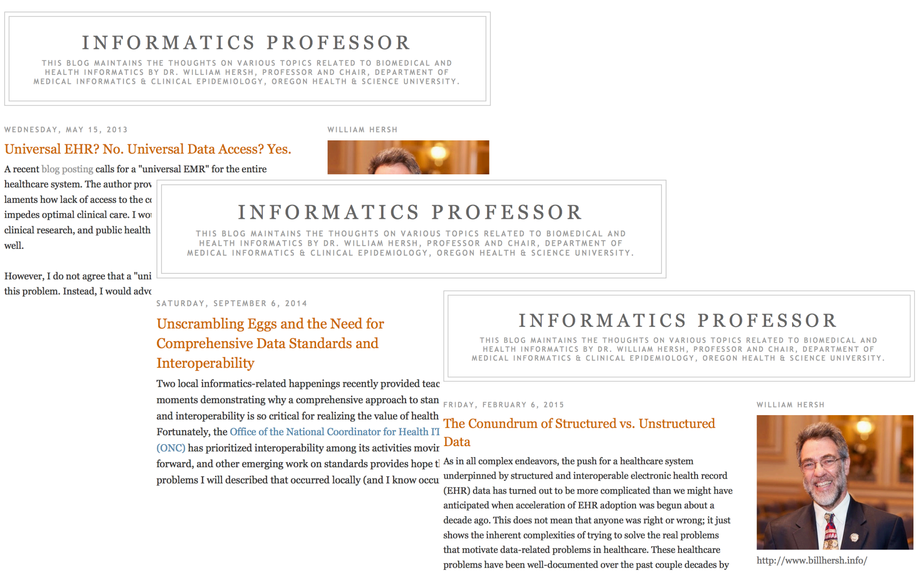
OREGON HEALTH & SCIENCE UNIVERSITY

"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

OREGON HEALTH & SCIENCE UNIVERSITY

A big challenge is interoperable data


INFORMATICS PROFESSOR

THIS BLOG MAINTAINS THE THOUGHTS ON VARIOUS TOPICS RELATED TO BIOMEDICAL AND HEALTH INFORMATICS BY DR. WILLIAM HERSH, PROFESSOR AND CHAIR, DEPARTMENT OF MEDICAL INFORMATICS & CLINICAL EPIDEMIOLOGY, OREGON HEALTH & SCIENCE UNIVERSITY.

WEDNESDAY, MAY 15, 2013

Universal EHR? No. Universal Data Access? Yes.

A recent blog posting calls for a "universal EMR" for the entire healthcare system. The author prov laments how lack of access to the c impedes optimal clinical care. I wo clinical research, and public health well.

However, I do not agree that a "uni this problem. Instead, I would adv

WILLIAM HERSH

INFORMATICS PROFESSOR

THIS BLOG MAINTAINS THE THOUGHTS ON VARIOUS TOPICS RELATED TO BIOMEDICAL AND HEALTH INFORMATICS BY DR. WILLIAM HERSH, PROFESSOR AND CHAIR, DEPARTMENT OF MEDICAL INFORMATICS & CLINICAL EPIDEMIOLOGY, OREGON HEALTH & SCIENCE UNIVERSITY.

SATURDAY, SEPTEMBER 6, 2014

Unscrambling Eggs and the Need for Comprehensive Data Standards and Interoperability

Two local informatics-related happenings recently provided teac moments demonstrating why a comprehensive approach to stan and interoperability is so critical for realizing the value of health Fortunately, the Office of the National Coordinator for Health IT (ONC) has prioritized interoperability among its activities movir forward, and other emerging work on standards provides hope t problems I will described that occurred locally (and I know occu

INFORMATICS PROFESSOR

THIS BLOG MAINTAINS THE THOUGHTS ON VARIOUS TOPICS RELATED TO BIOMEDICAL AND HEALTH INFORMATICS BY DR. WILLIAM HERSH, PROFESSOR AND CHAIR, DEPARTMENT OF MEDICAL INFORMATICS & CLINICAL EPIDEMIOLOGY, OREGON HEALTH & SCIENCE UNIVERSITY.

FRIDAY, FEBRUARY 6, 2015

The Conundrum of Structured vs. Unstructured Data

As in all complex endeavors, the push for a healthcare system underpinned by structured and interoperable electronic health record (EHR) data has turned out to be more complicated than we might have anticipated when acceleration of EHR adoption was begun about a decade ago. This does not mean that anyone was right or wrong; it just shows the inherent complexities of trying to solve the real problems that motivate data-related problems in healthcare. These healthcare problems have been well-documented over the past couple decades by

WILLIAM HERSH

http://www.billhersh.info/

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)
      – http://wiki.hl7.org/index.php?title=FHIR_for_Clinical_Users

OREGON HEALTH & SCIENCE UNIVERSITY

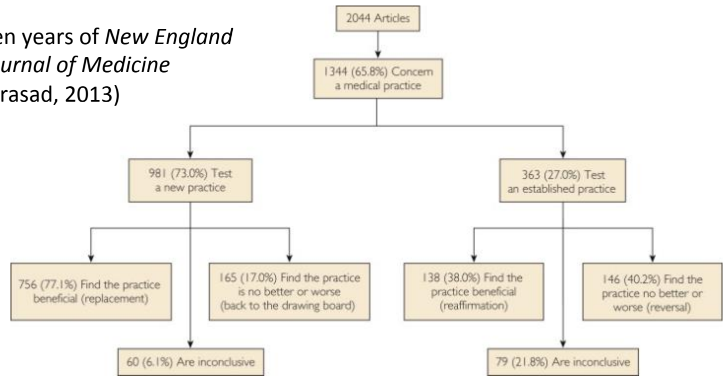
Also need to develop clinical data research networks

- Established
  - HMO Research Network – facilitates clinical research
    - www.hmoresearchnetwork.org
  - FDA Mini-Sentinel Network – safety surveillance
    - www.mini-sentinel.org
- New
  - PCORnet – 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

OREGON HEALTH & SCIENCE UNIVERSITY

Large amounts of data do not obviate the need for experimental research

Ten years of *New England Journal of Medicine* (Prasad, 2013)



- 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)

OREGON HEALTH & SCIENCE UNIVERSITY

23

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)

OREGON HEALTH & SCIENCE UNIVERSITY

24

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

Graduates	CI	BCB	HIM	Total
GC	321	0	37	358
MBI	146	6	2	154
MS	68	9	0	77
PhD	10	6	0	16
Total	545	21	39	605

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

27

Informatics Discovery Lab: academia-industry collaboration

http://www.ohsu.edu/idl

28

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-74duDDvwU
  - 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)
  - http://www.nlm.nih.gov