The Data-Rich, Information-Driven Future of Healthcare

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


http://www.biomedcentral.com/1472-6947/11/50


Roebuck, C (2012). The importance of knowing context of hospital episode statistics when reconfiguring the NHS. *British Medical Journal: Rapid Response.* [http://www.bmj.com/content/344/bmj.e2432/rr/578977](http://www.bmj.com/content/344/bmj.e2432/rr/578977)


The Data-Rich, Information-Driven Future of Healthcare

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

Outline

• Our dysfunctional healthcare system and a vision for fixing and optimizing it
• Part of the solution includes adoption of electronic health records (EHRs) and other technologies
• Toward the data-rich, information-driven learning health system
• Challenges in getting to such a system
• The skills and workforce need to get there
Our healthcare system is broken in many ways and needs fixin’

- Action must be taken to address (Smith, 2012)
  - $750B in waste (out of $2.5T system)
  - 75,000 premature deaths
- Sources of waste – from Berwick (2012)
  - Unnecessary services provided
  - Services inefficiently delivered
  - Prices too high relative to costs
  - Excess administrative costs
  - Missed opportunities for prevention
  - Fraud
- One vision for repair is the IOM’s “learning healthcare system” (Smith, 2012)

Recommendations for Best Care, Lower Cost (Smith, 2012)

- Foundational elements
  - Digital infrastructure
  - Data utility
- Care improvement targets
  - Clinical decision support
  - Patient-centered care
  - Community links
  - Care continuity
  - Optimized operations
- Supportive policy environment
  - Financial incentives
  - Performance transparency
  - Broad leadership

Health information technology (HIT) is part of solution

- Systematic reviews (Chaudhry, 2006; Goldzweig, 2009; Buntin, 2011) have identified benefits in a variety of areas
  - Although 18-25% of studies come from a small number of ‘health IT leader’ institutions
The world is adopting EHRs (Schoen, 2012)

Even in the US

(Hsaio, CDC, 2012)

(King, ONC, 2012)
Although advanced functionality is less common (Schoen, 2012)

Multifunctional health IT capacity – use of at least two electronic functions: order entry management, generating patient information, generating panel information, and routine clinical decision support

Why has it been so difficult to get there? (Hersh, 2004)

- Cost
- Technical challenges
- Interoperability
- Privacy and confidentiality
- Workforce
But now we have substantial investment in 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 $118M for workforce development)

Setting the stage for a data-rich, information-driven healthcare future

• Use of analytics and/or business intelligence (BI)
  – Analytics is the use of data collection and analysis to optimize decision-making (Davenport, 2010)
  – BI is the “processes and technologies used to obtain timely, valuable insights into business and clinical data” (Adams, 2011)

• As in many areas of advanced information and technology, healthcare and biomedicine are behind the curve of other industries (Miller, 2011; Rhoads, 2012)
Who employs analytics and BI outside of healthcare?

- Amazon and Netflix recommend books and movies with great precision
- Many sports teams, such as the Oakland Athletics and New England Patriots, have used “moneyball” to select players, plays, strategies, etc. (Lewis, 2004; Davenport, 2007)
- Facebook can target advertising (Ugander, 2011) and predict your location, political views, sexual preferences, and intelligence (Kosinski, 2013)
- Twitter volume and other linkages can predict stock market prices (Ruiz, 2012), though may be more effective at spreading misinformation (Salathe, 2013)
- Recent US election showed value of using data: re-election of President Obama (Scherer, 2012) and predictive ability of Nate Silver (Salant, 2012)

Levels of BI (Adams, 2011)

<table>
<thead>
<tr>
<th>BI Type</th>
<th>Example Uses</th>
<th>Questions Answered</th>
<th>BI Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization</td>
<td>Diagnostic and therapeutic approaches</td>
<td>How can we achieve the best outcome?</td>
<td>Prescriptive</td>
</tr>
<tr>
<td>Predictive modeling</td>
<td>Identify high-risk patients</td>
<td>What will happen next if...?</td>
<td>Predictive</td>
</tr>
<tr>
<td>Forecasting</td>
<td>Public health issues</td>
<td>What if these trends continue?</td>
<td>Predictive</td>
</tr>
<tr>
<td>Simulation</td>
<td>Business processes</td>
<td>What could happen if...?</td>
<td>Predictive</td>
</tr>
<tr>
<td>Alerts</td>
<td>Infection outbreaks</td>
<td>When are actions needed?</td>
<td>Descriptive</td>
</tr>
<tr>
<td>Query/drill-down</td>
<td>“Stop and drop”</td>
<td>What exactly is the problem?</td>
<td>Descriptive</td>
</tr>
<tr>
<td>Ad hoc reporting</td>
<td>Quick-change metrics</td>
<td>How many, how often, where?</td>
<td>Descriptive</td>
</tr>
<tr>
<td>Standard reporting</td>
<td>Key metrics</td>
<td>What happened?</td>
<td>Descriptive</td>
</tr>
</tbody>
</table>
Analytics part of larger “secondary use” or “re-use” of clinical data

- Many secondary uses or re-uses of electronic health record (EHR) data (Safran, 2007); these include
  - Using data to improve care delivery – predictive analytics
  - Healthcare quality measurement and improvement
  - Clinical and translational research – generating hypotheses and facilitating research
  - Public health surveillance – including for emerging threats
  - Implementing the learning health system

Using data analytics 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)
- Being applied by many now to problem of early hospital re-admission, in particular within 30 days (Sun, 2012; Gildersleeve, 2013)
- Prediction not only of patient response but also behavior, e.g., regimen adherence (Steffes, 2012)
- A requirement of coming “precision medicine” (Mirnezami, 2012)
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

<table>
<thead>
<tr>
<th>NQF Measure Number &amp; PQRI Implementation Number</th>
<th>Clinical Quality Measure Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>NQF 0013</td>
<td>Hypertension Blood Pressure Measurement</td>
</tr>
<tr>
<td>NQF 0028</td>
<td>Preventive Care and Screening Measure - a) Tobacco Use Assessment, b) Tobacco Cessation Intervention</td>
</tr>
<tr>
<td>NQF 0421 PQRI 128</td>
<td>Adult Weight Screening and Follow-up</td>
</tr>
</tbody>
</table>

Alternate Clinical Quality Measures

| NQF 0024                                      | Weight Assessment and Counseling for Children and Adolescents |
| NQF(041 PQRI 110)                             | Preventive Care and Screening: Influenza Immunization for Patients 50 Years Old or Older |
| NQF 0038                                      | Childhood Immunization Status |

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)
Clinical and translational research

• Led in part by activities of NIH Clinical and Translational Science Award (CTSA) Program (Mackenzie, 2012)
• eMERGE Network – connecting genotype-phenotype, http://emerge.mc.vanderbilt.edu
  – Has used EHR data to identify genomic variants associated with atrioventricular conduction abnormalities (Denny, 2010), red blood cell traits (Kullo, 2010), while blood cell count abnormalities (Crosslin, 2012), thyroid disorders (Denny, 2011), etc.
• Other successes include replication of RCTs findings of
  – Women’s Health Initiative (Tannen, 2007; Weiner, 2008)
  – Other cardiovascular diseases (Tannen, 2008; Tannen, 2009) and value of statin drugs in primary prevention of coronary heart disease (Danaei, 2011)

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)
• One notable success is Google Flu Trends (http://www.google.org/flutrends/) – search terms entered into Google predicted flu activity but not enough to intervene (Ginsberg, 2009); last year, performance was poorer (Butler, 2013)
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
Inaccurate

• Documentation not always a top priority for busy clinicians (de Lusignan, 2005)
• Data entry errors in a recent analysis in the English National Health Service (NHS) – yearly hospital statistics showed approximately (Brennan, 2012)
  – 20,000 adults attending pediatric outpatient services
  – 17,000 males admitted to obstetrical inpatient services – mainly due to male newborns (Roebuck, 2012)
  – 8,000 males admitted to gynecology inpatient services

Inaccurate (cont.)

• Analysis of EHR systems of four known national leaders assessed use of data for studies on treatment of hypertension and found five categories of reasons why data were problematic (Savitz, 2012)
  – Missing
  – Erroneous
  – Un-interpretable
  – Inconsistent
  – Inaccessible in text notes
Incomplete

• 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)
• Quality measures under-reported based on under-capture of data due to variation in clinical workflow and documentation practices (Parsons, 2012)
• Ability to identify diabetic patients successively increased as time frame of assessing records was increased from one through ten years of analysis (Wei, 2013)

Incomplete (cont.)

• Studies of health information exchange (HIE)
  – Study of 3.7 million patients in Massachusetts found 31% visited two or more hospitals over five years (57% of all visits) and 1% visited five or more hospitals (10% of all visits) (Bourgeois, 2010)
  – Analysis of 2.8 million emergency department patients in Indiana found 40% had data at multiple institutions (Finnell, 2011)
Unrecoverable for research

- Many clinical data are “locked” in narrative text reports (Hripcsak, 1995; Hripcsak, 2012), including summaries of care (D’Amore, 2012)
- A promising approach for recovering these data for research is natural language processing (NLP) (Nadkarni, 2011)
  - Has been most successful when applied to the determination of specific data elements, e.g., eMERGE studies (Denny, 2012)
- State of the art for performance of NLP has improved dramatically over the last couple decades, but is still far from perfect (Stanfill, 2010)
  - Still do not not know how good is “good enough” for NLP in data re-use for research, quality, etc. (Hersh, 2005)

Many data “idiosyncrasies”

- “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
Most important requirement may be changing healthcare system

- US healthcare system still mostly based on fee for service model – little incentive for managing care in coordinated manner
- Informatics tools can help in care coordination (Dorr, 2007; Dorr, 2008)
- Primary care medical home (PCMH) might be first step to improving value and providing incentive for better use of data (Longworth, 2011)
- Affordable Care Act (ACA) implements accountable care organizations (ACOs), which provide bundled, quality-adjusted payments for conditions (Longworth, 2011)
  - Being implemented statewide in Oregon Medicaid system (Stecker, 2013)

Another need is for academic programs leading in research and education

www.ohsu.edu/informatics
www.ohsu.edu/idl
DMICE/IDL faculty areas of expertise

**Expertise**

- Healthcare analytics
- Workflow redesign
- Qualitative and quantitative evaluation methodologies
- Natural language processing, machine learning, and information retrieval
- Biomedical terminologies, ontologies, and coding
- Biomedical data structures, representation, and normalization

**Focus Areas**

- Predictive Analytics for Healthcare Process Redesign
- Usability for EHR Data and Users
- Population Management and Care Coordination Information Systems
- Mobile Telemedicine
- Precision Medicine

**Conclusions**

- A growing body of evidence supports EHR and other IT to improve health and healthcare
- The world is gradually adopting EHRs and other IT
- The next step is to make use of the increasing data through analytics and BI to achieve the learning healthcare system
- There are challenges, but also benefits, to this use data-driven, information-driven evolution
- There is a growing need for research and education in all areas of informatics
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/whats
• Office of the National Coordinator for Health IT (ONC)
  – http://www.healthit.gov
• American Medical Informatics Association (AMIA)
  – http://www.amia.org
• National Library of Medicine (NLM)