Healthcare Data Analytics

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
Web: www.billhersh.info
Blog: http://informaticsprofessor.blogspot.com

References


Kellermann, AL and Jones, SS (2013). What will it take to achieve the as-yet-unfulfilled promises of health information technology? Health Affairs. 32: 63-68.


[http://stm.sciencemag.org/content/4/158/158rv11.short](http://stm.sciencemag.org/content/4/158/158rv11.short)


Healthcare Data Analytics

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
Web: www.billhersh.info
Blog: http://informaticsprofessor.blogspot.com

Healthcare data analytics

• Rationale
• Definitions
• Applications
• Results
• Challenges
• Workforce
• Further study
Rationale

- Although focus in recent years has been on EHR implementation and “capture/share data” of Stage 1 meaningful use (MU), informatics work in the future will shift to putting the data and information to good use (Hersh, 2012)
- As the quantity and complexity of healthcare data grow through EHR capture, genomics, and other sources, the number of facts per clinical decision will increase, requiring increasing help for decision-makers (Stead, 2011)

Definitions

- Both a buzz-word and an important emerging area
- Davenport (2007) – “the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions”
- IBM (2012) – “the systematic use of data and related business insights developed through applied analytical disciplines (e.g. statistical, contextual, quantitative, predictive, cognitive, other [including emerging] models) to drive fact-based decision making for planning, management, measurement and learning”
Levels of analytics (Adams, 2011)

<table>
<thead>
<tr>
<th>Degree of Competitive Advantage and Complexity</th>
<th>Diagnostic and Therapeutic Approaches</th>
<th>Prescriptive</th>
<th>Predictive</th>
<th>Descriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization</td>
<td>How can we achieve the best outcomes?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictive modeling</td>
<td>Identify high-risk patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasting</td>
<td>Public health issues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td>Business processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alerts</td>
<td>Infection outbreaks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Query/drill-down</td>
<td>&quot;What and why?&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ad hoc reporting</td>
<td>Out-of-range metrics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard reporting</td>
<td>Key metrics</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Related terms

- Machine learning – area of computer science focused on systems and algorithms that learn from data (Flach, 2012; Crown, 2015)
- Data mining – processing and modeling of data to discover previously unknown patterns or relationships (Bellazzi, 2008; Zaki, 2014)
- Text mining – applying data mining to unstructured textual data (Aggarwal, 2012)
- Big data – data of growing volume, velocity, variety, and veracity (Zikopolous, 2011; O’Reilly, 2015)
  - e.g., ~9 petabytes of data of Kaiser-Permanente (Gardner, 2013)
Related terms (cont.)

- Data science – distinguished from statistics by understanding of varying types and how to manipulate and leverage (Dhar, 2013; Grus, 2015)
- Data provenance – origin and trustworthiness (Buneman, 2010)
- Business intelligence – use of data to obtain timely, valuable insights into business and clinical data (Adams, 2011)
- Personalized (Hamburg, 2010), precision (IOM, 2011; Collins, 2015; Ashley, 2015), or computational medicine (Winslow, 2012)

Analytics pipeline

- Adapted from Kumar (2013) for healthcare (Hersh, 2014)
Analytics is well-employed 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)
• Twitter volume and other linkages can predict stock market prices (Ruiz, 2012)
• US 2012 election showed value of using data: re-election of President Obama (Scherer, 2012) and predictive ability of Nate Silver (Salant, 2012)
• Individual traits such as sexual orientation, political affiliation, personality types, and ethnicity can be discerned from Facebook “likes” with high accuracy (Kosinski, 2013)
• “Internet advertising” is a growing area (Smith, 2014), aiming to solve “Wanamaker dilemma” (O’Reilly, 2012)
• Government (e.g., National Security Agency in US) tracking of email, phone calls, and other digital trails (Levy, 2014)

What about analytics in healthcare?

• With shift of payment from “volume to value,” healthcare organizations will need to manage information better to deliver better care (Diamond, 2009; Horner, 2012)
  – To realize this, they must achieve “analytic integration” (Davenport, 2012)
• New care delivery models (e.g., accountable care organizations) will require better access to data (e.g., health information exchange, HIE)
  – Halamka (2013): ACO = HIE + analytics
• Recent overviews (Burke, 2013; Gensinger, 2014; Marconi, 2014)
Applications of analytics in healthcare

- Early application – identifying patients at risk for hospital readmission within 30 days of discharge
- Centers for Medicare and Medicaid Services (CMS) Readmissions Reduction Program penalizes hospitals for excessive numbers of readmissions (2013)
- Several studies have used EHR data to predict patients at risk for readmission (Amarasingham, 2010; Donzé, 2013; Gildersleeve, 2013; Shadmi, 2015)

Applications of analytics – identifying other clinical situations

- Predicting 30-day risk of readmission and death among HIV-infected inpatients (Nijhawan, 2012)
- Identification of children with asthma (Afzal, 2013)
- Detecting postoperative complications (FitzHenry, 2013)
- Measuring processes of care (Tai-Seale, 2013)
- Determining five-year life expectancy (Mathias, 2013)
- Detecting potential delays in cancer diagnosis (Murphy, 2014)
- Identifying patients with cirrhosis at high risk for readmission (Singal, 2013)
- Predicting out of intensive care unit cardiopulmonary arrest or death (Alvarez, 2013)
- Predicting hospital death by day or time of day (Coiera, 2014)
- Predicting future patient costs (Charlson, 2014)
Applications of analytics – patient identification and diagnosis

• Identifying patients who might be eligible for participation in clinical studies (Voorhees, 2012)
• Determining eligibility for clinical trials (Köpcke, 2013)
• Identifying patients with diabetes and the earliest date of diagnosis (Makam, 2013)
• Predicting diagnosis in new patients (Gottlieb, 2013)

Most important use cases for data analytics (Bates, 2014)

• High-cost patients – looking for ways to intervene early
• Readmissions – preventing
• Triage – appropriate level of care
• Decompensation – when patient’s condition worsens
• Adverse events – awareness
• Treatment optimization – especially for diseases affecting multiple organ systems
Requirements for data analytics in healthcare

- Infrastructure (Amarasingham, 2014)
  - Stakeholder engagement
  - Human subjects research protection
  - Protection of patient privacy
  - Data assurance and quality
  - Interoperability of health information systems
  - Transparency
  - Sustainability
- New models of thinking and training (Krumholz, 2014)
- New tools, e.g., “green button” to help clinicians aggregate data in local EHR (Longhurst, 2014)

Results of analytics in improving patient outcomes

- Readmission tool applied to case management approach helped reduce readmissions (Gilbert, 2013)
- Bayesian network model embedded in EHR to predict hospital-acquired pressure ulcers led to tenfold reduction in ulcers and one-third reduction in intensive care unit length of stay (Cho, 2013)
- Readmission risk tool intervention reduced risk of readmission for patients with congestive heart failure but not those with acute myocardial infarction or pneumonia (Amarasingham, 2013)
- Automated prediction model integrated into existing EHR successfully identified patients on admission who were at risk for readmission within 30 days of discharge but had no effect on 30-day all-cause and 7-day unplanned readmission rates over 12 months (Baillie, 2013)
Challenges for analytical use of clinical data

- Data quality and accuracy is not a top priority for busy clinicians (de Lusignan, 2005)
- Patients get care at different places (Bourgeois, 2010; Finnell, 2011)
- Standards and interoperability – mature approaches but lack of widespread adoption (Kellermann, 2013)
- Much data is “locked” in text (Hripcsak, 2012)
- Average pediatric ICU patient generates 1348 information items per 24 hours (Manor-Shulman, 2008)

Caveats for use of operational EHR data (Hersh, 2013) – may be

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

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

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apply an Evidence-Based Approach</td>
<td>Ask an answerable question, find the best EHR data (&quot;evidence&quot;), appraise the data, apply evidence to question</td>
</tr>
<tr>
<td>Evaluate and Manage Data</td>
<td>Assess availability, completeness, quality (validity), and transformability of data</td>
</tr>
<tr>
<td>Create Tools for Data Management</td>
<td>Create software (especially pipelines) for data aggregation, validation and transformation</td>
</tr>
<tr>
<td>Determine Metrics for Data Assessment</td>
<td>Determine whether a particular site’s data are &quot;research grade&quot;</td>
</tr>
<tr>
<td>Develop Methods for Comparative Validation</td>
<td>Develop tools that support analysis of multi-site data collections</td>
</tr>
<tr>
<td>Develop a Methodology Knowledge Base</td>
<td>Develop a data catalogue that relates data elements to recommended transformations</td>
</tr>
<tr>
<td>Standardize Reporting Methods</td>
<td>Provide details of data sources, provenance and manipulation, to support comparison of data</td>
</tr>
<tr>
<td>Engage Informatics Expertise</td>
<td>Ensure validity of findings derived from data collected from disparate sources</td>
</tr>
<tr>
<td>Include an Informatics Research Agenda</td>
<td>Generate systematic studies of inherent biases in EHR and data collection methods, such as data entry user interfaces</td>
</tr>
</tbody>
</table>
Apply an evidence-based medicine approach (Hersh, 2013)?

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

Analytics workforce

- Data scientists – the “sexiest profession of the 21st century” (Davenport, 2012)
- Key skill sets include
  - Machine learning, based upon a foundation of statistics (especially Bayesian), computer science (representation and manipulation of data), and knowledge of correlation and causation (modeling) (Dhar, 2013)
  - IBM – both “numerate” and business-oriented skills (Fraser, 2013)
  - NIH – big data researchers need training in quantitative sciences, domain expertise, ability to work in diverse teams, and understanding concepts of managing and sharing data (NIH, 2013)
How many are needed?

- McKinsey (Manyika, 2011) – need in US in all industries (not just healthcare) for
  - 140,000-190,000 individuals who have “deep analytical talent”
  - 1.5 million “data-savvy managers needed to take full advantage of big data”

- In UK, estimated by 2018 will be over 6400 organizations that will hire 100 or more analytics staff (SAS, 2013)

What skills are needed (Hersh, 2014)?

- Programming – especially with data-oriented tools, such as SQL and statistical packages
- Statistics – working knowledge to apply tools and techniques
- Domain knowledge
- Communication – ability to understand needs of people and organizations and articulate results back to them

- Is this informatics? Or a specialization of informatics? Or something totally different?
How can I learn more in Oregon?
Study informatics?

- Many educational opportunities at a variety of levels, mostly graduate
  - [http://www.amia.org/informatics-academic-training-programs](http://www.amia.org/informatics-academic-training-programs)
- OHSU program one of largest and well-established (Hersh, 2007)
  - [http://www.ohsu.edu/informatics-education](http://www.ohsu.edu/informatics-education)
  - Graduate level programs at Certificate, Master’s, and PhD levels
  - “Building block” approach allows courses to be carried forward to higher levels

OHSU program has three tracks

- Clinical Informatics
  - Original track, focused on informatics in health, healthcare, public health, and clinical research settings
- Bioinformatics and Computational Biology (BCB)
  - Focused on informatics in genomics, molecular biology, and their translational research aspects
- Health Information Management (HIM)
  - Overlapping with clinical informatics, focused on HIM profession and leading to Registered Health Information Administrator (RHIA) certification
OHSU offers a variety of degrees and certificates

- Doctor of Philosophy (PhD)
  - For those who wish to pursue research, academia, or leadership careers
- Master of Science (MS)
  - Research master’s, including for those with doctoral degrees in other fields who wish to pursue research careers
- Master of Biomedical Informatics (MBI)
  - Professional master’s degree for practitioners and leaders
- Graduate Certificate
  - Subset of master’s degree as an introduction or career specialization

<table>
<thead>
<tr>
<th>Degree/Certificate Track</th>
<th>PhD</th>
<th>MS</th>
<th>MBI</th>
<th>Grad Cert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Informatics</td>
<td>On-campus</td>
<td>On-campus</td>
<td>On-campus</td>
<td>On-campus</td>
</tr>
<tr>
<td>Bioinformatics and Computational Biology</td>
<td>On-campus</td>
<td>On-campus</td>
<td>On-campus</td>
<td>On-campus</td>
</tr>
<tr>
<td>Health Information Management</td>
<td>On-campus</td>
<td>On-campus</td>
<td>On-campus</td>
<td>On-campus</td>
</tr>
<tr>
<td></td>
<td>On-line</td>
<td>On-line</td>
<td>On-line</td>
<td>On-line</td>
</tr>
</tbody>
</table>
Overview of OHSU graduate programs

<table>
<thead>
<tr>
<th>Masters</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Tracks:</td>
</tr>
<tr>
<td>- Clinical Informatics</td>
</tr>
<tr>
<td>- Bioinformatics</td>
</tr>
<tr>
<td>- Thesis or Capstone</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Graduate Certificate</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Tracks:</td>
</tr>
<tr>
<td>- Clinical Informatics</td>
</tr>
<tr>
<td>- Health Information Management</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10x10</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Or introductory course</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PhD</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Knowledge Base</td>
</tr>
<tr>
<td>- Advanced Research Methods</td>
</tr>
<tr>
<td>- Biostatistics</td>
</tr>
<tr>
<td>- Cognate</td>
</tr>
<tr>
<td>- Advanced Topics</td>
</tr>
<tr>
<td>- Doctoral Symposium</td>
</tr>
<tr>
<td>- Mentored Teaching</td>
</tr>
<tr>
<td>- Dissertation</td>
</tr>
</tbody>
</table>

Conclusions

- There are plentiful opportunities for data analytics in healthcare
- We must be cognizant of caveats of using operational clinical data
- We must implement best practices for using such data
- There are also opportunities for HIM and informatics professionals in healthcare data analytics
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://www.healthit.gov
- American Medical Informatics Association (AMIA)
  - http://www.amia.org
- National Library of Medicine (NLM)