Big Data Science and Analytics in Health and Biomedicine

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


http://www.sei.cmu.edu/measurement/research/upload/Davidson.pdf


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
Big Data Science and Analytics in Health and Biomedicine

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

Big Data Science and Analytics in Healthcare

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

• Although focus in recent years has been on electronic health record (EHR) implementation and “meaningful use,” 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>How can we achieve the best outcomes?</th>
<th>Prescriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization</td>
<td>Identify high-risk patients</td>
<td>What will happen next if...?</td>
<td>Predictive</td>
</tr>
<tr>
<td>Predictive modeling</td>
<td>Public health issues</td>
<td>What if these trends continue?</td>
<td>Descriptive</td>
</tr>
<tr>
<td>Forecasting</td>
<td>Business processes</td>
<td>What could happen if...?</td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alerts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Query/drill-down</td>
<td>“Silo and drill”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ad hoc reporting</td>
<td>Output-range metrics</td>
<td>How many, how often, where?</td>
<td></td>
</tr>
<tr>
<td>Standard reporting</td>
<td>Key metrics</td>
<td>What happened?</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)
• 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)

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)
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

• OHSU program one of largest and well-established (Hersh, 2007)
  – 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

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 many career opportunities in healthcare data analytics