Big Data Analytics in the Clinical Laboratory

Nevada Health Care Forum
October 4th, 2016

Noah Traisman
Clinical Informatics Executive
Quest Quanum - Information Ventures
Objectives

• Define big data

• Learn how laboratories have used laboratory data analytics to:
  • Improve quality
  • Serve health plans/systems and deliver public health reports
  • Provide new tests, such as next-generation sequencing-based assays in the fields of oncology and neurology
  • Create public health insights

• Define the new tools of the trade

• Explore how big data may change healthcare
Definition of big data

• Volume
• Velocity
• Variety
• Variability
• Complexity
Garbage in, insights out

- Much of the data we have is unconnected to relevant information
  - Medicare payment data is already available. It says nothing about value. It is fraught with potential misinterpretations. What are the measures or proxies for quality?

- How do we relate observations with outcomes? Which are associations are coincidental and which are causal?
  - Is there a relationship between reduced movement and cholera outbreaks? Or is the relationship coincidental: flooding causes both reduced movement and reduced sanitation quality, the latter of which leads to cholera outbreaks.

- Big data doesn’t ask “why?” Big data detects new patterns.
  - Customers who buy anti-scuff furniture polish are less likely to default on home loans. Maybe people who take better care of their furniture are more likely to plan for the long-term and take pride in their possessions.
  - Patients with multiple addresses in the past year have higher hospital re-admission rates. Possibly, these patients have less social support upon discharge.
How laboratories have used big laboratory data analytics

- Test results
- Proficiency testing
- Quality control (QC)
- Inter-laboratory quality control
- Next-generation sequencing (NGS), oncology
- Health plan and public health reporting
- Aggregated laboratory data (public health insights)
- Aggregated patient data (patient care insights)
Proficiency testing

- The ability to compare results from different laboratories; define acceptable and unacceptable performance
- Develop statistical applications to better evaluate and monitor method systems to assure reliability of test results
Quality control

- Laboratories have been running quality control for many decades
- Walter Shewhart (Western Electric Company, Hawthorne Works)
  - Assignable-cause versus chance-cause
- Levey-Jennings charts
- Westgard

Inter-laboratory quality control

Some manufacturers of quality control material began offering programs where participating laboratories could compare their own results to those of other laboratories.

Networks of laboratories can compare performance across all laboratories in the network.
Next-generation sequencing

The sophistication needed to analyze genetic data required new tools to interpret and control.

- Analytic validity
- Clinical validity
- Clinical utility

Health plan and public health reporting

- Mapping disease
- Developing standards
- Recognizing performance
- Paying for better outcomes
Aggregated laboratory data

- Quest Diagnostics Health Trends™
  - H1N1, Rotavirus, HIV
  - Diabetes, chronic kidney disease
  - Impact of the Affordable Care Act
  - Gaps in care in pregnancy and post-partum
  - Vitamin D and parathyroid hormone
  - Allergies Across America™
  - Newly identified conditions in employer wellness program
  - Warfarin in patients with atrial fibrillation
  - Misuse of drugs, prescription drug monitoring
  - Lead
  - Cervical disease testing
  - LDL cholesterol
Quest Diagnostics reviewed more than 125 million urine drug tests administered from 1988 to last year. Overall, 3.5% of specimens came back positive (in 2012) compared with 13.6% in 1988. The vast majority of tests, around 75% in recent years, were conducted for pre-employment screening. The rest were administered following accidents, after employers suspected drug use or as part of regular testing regimens.”

*Wall Street Journal, November 18, 2013*
I want my healthcare data to be available where and when it is needed to make the right healthcare decisions for routine and emergency use.

- Contact information for my personal physician and dentist
- Eye prescription
- Emergency contact information
- Copy of health directive
- Medical and family history

Electronic Patient Health Record
Big data in healthcare

- Massive data sharing and analysis in healthcare are lagging retail and banking
- CMS is a leader in healthcare with a focus on fraud detection
Big data in healthcare: big data tools

- Apache Hadoop, open-sourced, developed by Yahoo
  - Microsoft’s Polybase is a query tool that enables use of Hadoop Distributed File Systems and SQL relational databases
  - Impala enables the use of SQL over Hadoop databases
  - A Hadoop cluster is typically built from inexpensive, commodity hardware and typically runs on traditional disk drives rather than in expensive storage area networks
  - Enterprise data warehouse is used to store structured and unstructured data
- Query languages are complex. MapReduce algorithms are sophisticated
- Big data is inexpensive but hard to use
Big data in healthcare

- National Cancer Institute’s Surveillance, Epidemiology and End Results (SEER) and SEER-Medicare
  - Demographics (age, gender, ethnicity, marital status)
  - Diagnosis (origin, stage, histology, treatment, survival)
  - SEER-Medicare contains additional information provided to Medicare recipients

- American College of Surgeons’ National Cancer Database

<table>
<thead>
<tr>
<th></th>
<th>Big Data</th>
<th>Clinical Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounding</td>
<td></td>
<td>Advantage</td>
</tr>
<tr>
<td>Detailed data elements</td>
<td></td>
<td>Advantage</td>
</tr>
<tr>
<td>Sample size</td>
<td>Advantage</td>
<td></td>
</tr>
<tr>
<td>Timely results</td>
<td>Advantage</td>
<td></td>
</tr>
<tr>
<td>Generalizability</td>
<td>Advantage</td>
<td></td>
</tr>
</tbody>
</table>
Big data in healthcare

- The University of Ontario’s Institute of Technology, with IBM, was able to predict nosocomial infections in the neonatal intensive care unit 24 hours before symptoms developed.

- Dr. Atul Gawande and Brigham and Women’s Hospital found by systematically standardizing knee joint-replacement surgery they could improve outcomes and reduce costs.

- Kaiser Permanente discovered adverse drug events with Vioxx® that led to the drug being withdrawn.

- Researchers from Johns Hopkins used Google Trends to identify early signals of the flu. Twitter updates were used to identify outbreaks of cholera two weeks earlier than otherwise in Haiti after the 2010 earthquake.

- Clinical decision support reduced adverse drug reactions in a pediatric critical care unit by 40% within two months.
Big data in healthcare

Shah ND, Pathak J. Why Health Care May Finally Be Ready for Big Data.

- Integrating data
- Generating knowledge
- Translating knowledge into practice

“Precision data diagnostics”
Big data in healthcare: clinical trials

Challenges

- Clinical trials enrollment of patients/subjects
- Clinical trials enrollment of investigators
- Monitoring progress of clinical trials
- Post-launch drug monitoring
Big data in healthcare

- The Health and Human Services’ Federal Health IT Strategic Plan is an attempt to coordinate efforts across federal agencies. A roadmap established these goals:
  - Establishing a coordinated governance framework and process for nationwide health IT interoperability
  - Improving technical standards and implementation guidance for sharing and using a common clinical database
  - Enhancing incentives for sharing electronic health information according to common technical standards
  - Clarifying privacy and security requirements that enable interoperability
Big data in healthcare: challenges

- Data reliability
  - How reliable is the primary data?
- Data elements
  - Develop standard terms and ontologies
  - Define elements in EHR
  - Invest in usability within clinical workflow
- Data security
  - Who has access?
  - Change in control
  - How will it be used?
  - Disposal/destruction
- Timeliness
  - Data must be available to act upon
- Test impact
  - Is model confirmed in practice?
- Conflicting models
  - Will there be competing conclusions?
- Who will use the information?
  - Information overload
- Governance
  - Distrust of big data
- Letting go of paper
Big data in healthcare

- Who should pay?
- Who will benefit?
Big data, big brother?

- We live in the world of “the internet of things”
- We live with tracking devices that monitor our activity, sleep, and stress
- How would decisions be made differently if we knew how well everyone slept the previous night, the stress levels of others involved in the decision, their activity levels in the previous day, and what they ate, drank, or inhaled? Would you want to know this about your surgeon and the operating room staff? How about the pathologist about to look at the slide of your tissue? How do my rights to know balance privacy rights?