Data as an Asset
Session 285, February 14, 2019
Ferdinand Velasco, MD, CHIO  Geoff Clark, Solution Principal
Conflict of Interest

Ferdinand Velasco, M.D.
Geoff Clark

Have no real or apparent conflicts of interest to report.
Agenda

• Learning objectives
• Big data in healthcare
• How to be data driven
• Case study: sepsis clinical pathway
Learning Objectives

• Recognize foundational data analytics methods and concepts and determine how to appropriately apply them to meet organizational needs

• Employ the communication skills required to effectively tell a story using data

• Identify the attributes of an effective data visualization for communicating to stakeholders
Big Data

A term used to refer to the study and applications of data sets that are so big and complex that traditional data-processing application software are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. There are a number of concepts associated with big data: originally there were 3 concepts volume, variety, velocity. Other concepts later attributed with big data are veracity (i.e., how much noise is in the data) and value. [Wikipedia]
Big Data in Healthcare

- Digitization of healthcare delivery
- Convergence of stakeholders
- Consumer orientation
On the Prospects for a (Deep) Learning Health Care System

C. David Naylor, MD, DPhil
Department of Medicine, University of Toronto, Ontario, Canada.

In 1976, Maxmen predicted that artificial intelligence (AI) in the 21st century would usher in “the post-physician era,” with health care provided by paramedics and computers. Today, the mass extinction of physicians remains unlikely. However, as outlined by Hinton in a related Viewpoint, the emergence of a radically different approach to AI, called deep learning, has the potential to effect major changes in clinical medicine and health care delivery. This Viewpoint reviews some of the factors driving wide adoption of deep learning and other forms of machine learning in the health ecosystem.

Emergence of Deep Learning
Forms of machine learning (eg, linear and logistic regression models) have been applied for decades in basic and applied research. However, recent advances in computing power and data availability have enabled the development of deep learning algorithms that can automatically extract complex patterns from large and high-dimensional datasets. These algorithms have shown remarkable success in various domains, including image and speech recognition, and have the potential to revolutionize health care.

University of Toronto was used to significantly improve speech recognition on Android devices and the prediction of drug activity in a Merck competition. It also substantially outperformed other approaches to computer vision in a public competition.

The technology quickly found wide commercial use. Early adopters included Google, Facebook, Microsoft, Apple, and Amazon, with the result that deep learning became globally ubiquitous almost overnight.

Deep learning had intuitive appeal for health-related applications, given its demonstrable strengths in intricate pattern recognition and predictive model building from high-dimensional data sets. These analytic capabilities have already proven useful for basic and applied researchers, ranging across health disciplines.

Big Data in Healthcare

- Invest: in modern business intelligence tools
- Develop: advanced analytics expertise
- Partner and collaborate: academia and private firms
- Influence culture: become a data driven organization
Becoming data-driven

- Be data-literate
- Be curious
- Be skeptical
- Be communicative
- Be action-oriented

Source: The Advisory Board Company
Be data-literate

- Understand the data (environment, source systems, workflows)
- Be savvy about performance measurement
- Know which analytical methods to use for which applications
- Seek opportunities for continuing education
Teach statistics before calculus!
Be curious

• Ask lots of questions
• Be open to learning novel subject areas
• Devote time for data exploration
Be skeptical

- Question reliability of source data
- Challenge conclusions
- Look for bias
- Discern whether the results are statistically significant and meaningful
Be communicative

- Tell a story using data
- Be effective at leveraging data visualization
- Understand audience
- Prepare well
Cole Nussbaumer Knaflic

storytelling with data

A data visualization guide for business professionals

WILEY
Data visualization strategies

• Consider the audience
• Select the appropriate visualization
  – Leverage the benefits of bar graphs
  – Avoid pie (and similar) charts
• Minimize clutter
• Utilize color strategically
• Optimize the use of text
Example

Application Assessment Review

31.5% is the percentage decrease in the total number of application assessment gaps from June to October due to the proactive management of potential harm to the organization indicated by the graph to the left. When applications are assessed, gaps are generated and placed into either a Managed or a Carried state. Managed gaps are further classified as:

- Policy Exceptions
- Assigned to the Risk Register
- Processed through Issue Management
Be action-oriented

- Focus on data analytics that matter
- Use data to improve
- Be agile
- Align with strategic priorities and key initiatives
- Embed predictive analytics into the workflow
Clinical pathway for sepsis

Case study
Sepsis

“...at its inception, is difficult to recognize but easy to treat; left unattended it becomes easy to recognize and difficult to treat.”

Machiavelli, The Prince (1513)
Sepsis

“…is a life threatening organ dysfunction caused by a dysregulated host response to infection. Sepsis is a global public health emergency, affecting millions of people worldwide, and representing one of the largest causes of death across the world.”

Surviving sepsis campaign: research priorities for sepsis and septic shock, Critical Care Medicine (2018)
What is a clinical pathway?

A multidisciplinary plan of care translating evidence-based medicine into discrete, timebound tasks or objectives.
The Problem and the Approach

• Not realizing the expected benefits of sepsis clinical pathway in terms of improved outcomes
• Internal analytics efforts yielding limited insights
• Engage third party data science consultants to leverage advanced data analytics techniques (8 week engagement)
  – Itemize and prioritize analytics questions
  – Data extracts/cleansing
  – Daily and weekly huddles
  – Review of results with executive sponsors
What kind of problem are we dealing with?

An entity engages with multiple parts of a complex organization, where interaction points are governed by defined protocols and engagement quality can be defined by a singular outcome measure.
Why is the problem hard to solve?

Process complexity, voluminous data and challenging signal detection make pathway evaluation hard.
Why is the problem hard to solve?

The processes are complex

Clinical pathways are often highly branched with downstream protocols contingent on earlier steps or patient responses. This can create a proliferation of sparsely populated measures and complicates the definition of compliance with the proposed pathway.
Why is the problem hard to solve?

The data is big

While EHR systems have improved the quality and accessibility of clinical data, the multidisciplinary nature of clinical pathways can still present data challenges. These can include multiple source systems, level of collection and aggregation, form (text, image, tabular) in addition to sheer size.
Why is the problem hard to solve?

Differentiating signal from noise is hard

Unaccounted for variables (e.g., comorbidity risk), error-filled data collection processes and noisy data hide the impact of clinical pathways on patient outcomes. A careful approach and a robust analytic methodology are required to differentiate signal from noise.
Traditional approaches and where they fail

**Simple Trending**
*Time series of patient outcome measures*

- **Pros**...
  - Easy to operationalize

- **Cons**...
  - Unless the signal is huge, it will be drowned by the noise
  - No visibility to ‘why’ the pathway is or isn’t working

**Group Differences**
*Separate patients by compliance and compare*

- **Pros**...
  - More informative than simple trending

- **Cons**...
  - Defining pathway compliance is problematic
  - No way to explicitly account for ‘hidden’ variables driving group differences

**Statistical Analysis**
*Measure statistical relationship between outcome and compliance measures*

- **Pros**...
  - Begins to quantify the relationship between outcomes and pathway variables

- **Cons**...
  - Stringent modeling assumptions are often violated
  - Pathway measures are often sparsely populated
Machine learning for Evaluating Pathways

Engineer
- Translate the pathway into data
- Model pathway externalities

Classify/Regress
- Frame the learning problem
- Choose models most suited to the data

Interrogate
- Let the models compete
- Quantify the relationships
- Ask the model what is important
- Triangulate among competing models
Sepsis Pathway Evaluation

CMS Core Measure Validation
• Do CMS guidelines optimize outcomes for the sepsis patient population?

Fluid for the Fluid Vulnerable
• What does the data say about fluid resuscitation for patients where aggressive fluid infusion may be problematic?

Sepsis and Flu Season
• Does flu season have a negative impact on sepsis patients and what should we do about it?
How Important is Timing of Fluid Administration?

Administering a fluid bolus within 3 hours of Severe Sepsis presentation is a core pillar of the CMS 3-hour bundle. But what is the specific relationship between bolus timing and mortality risk?

**Actuals, Predicted & Partial Dependence**

Extreme values obscure the subtle relationship found in the right hand part of the graph—where most of the data lies.
How Important is Timing of Fluid Administration?

Magnified Partial Dependence

Magnifying the region of interest reveals that mortality risk increases markedly at hour 1 and plateaus by hour 3. Earlier is definitely better for fluid administration in Sepsis.

Partial Dependence of Bolus Administration Timing on Sepsis Mortality
Fluid for the Fluid Vulnerable

- Fluid vulnerable patients are those with comorbid conditions that complicate aggressive fluid resuscitation (e.g., end stage renal failure, heart failure, pulmonary hypertension and obesity).

- Where fluid resuscitation is indicated, risk declines by 1.7% for optimal fluid volume (30 ml/kg), and risk increases only slightly for greater than optimal volumes.

Partial Dependence of Bolus Administration Volume on Sepsis Mortality
The Case for Repeated Lactic Acid Measures

Measuring lactate levels quantifies risk for patients with suspected sepsis and helps monitor the impact of therapy, but the cost of repeated lactates adds up.

The subsequent charts represent a scenario where Lactic Acid measures are repeated only when initial values are greater than 2.

Is there an impact to patient outcomes with this policy?

Partial Dependence of Initial Lactic Acid Value on Sepsis Mortality

Variable Importance Plot for Sepsis Mortality Classifier
The Case for Repeated Lactic Acid Measures

1. Dip in mortality risk
   This dip likely occurs because patients with initial lactates above 2.0 will get subsequent lactic acid measurements. The visibility and potential intervention enabled by repeat lactic acid measurements appears to be beneficial.

2. Second Lactate Importance
   In the classifier variable importance plot, second lactate value is ~27% more important than initial lactate value—demonstrating the usefulness of repeated lactate measures.
It’s been hypothesized that flu season adversely affects sepsis patients as Emergency Departments (ED) become clogged with Flu patients.

Here we show the relationship between ED load and Sepsis mortality for ‘Fluid Vulnerable’ patients. Flu season does seem to impact certain categories of Sepsis patients due to the burden placed on ED resources.
Sepsis and Flu Season

Partial Dependence of ED Total Load on Sepsis Mortality

ED Load From All Encounters
The graph shows Sepsis mortality risk as a function of percent load on ED. When ED loads approach historic highs, Sepsis patients see a step change in mortality risk.

Partial Dependence of ED Load from Flu on Sepsis Mortality

ED Load From Flu-like Encounters
The graph shows Sepsis mortality risk as a function of percent load on ED from patients with flu-like symptoms. When flu ED loads approach 5% of historic highs, Sepsis patients see a step change in mortality risk.
Outcomes

• Broad communications to stakeholder groups (executive and clinical leadership)

• Renewed focus on clinical pathway specifications shown to improve sepsis mortality
  – Testimonial: “This will change the way I practice medicine.” (ED physician)

• Implementation of predictive analytics to anticipate impact of flu outbreak
Outcomes

- Reduction in sepsis mortality
Key success factors

• Strong executive support
• Collaboration and customer engagement
• Time-constrained, disciplined approach: focus, prioritization, daily and weekly huddles
• Robust, reliable data source for analytics
• Broad exposure of findings to key stakeholder audiences
• Action orientation: reinforcement of interventions shown to improve outcomes
Questions

- Ferdinand Velasco, M.D.
  FerdinandVelasco@texashealth.org

- Geoff Clark
  geoff.clark@slalom.com