Big Data: New and Emerging Big Data Strategies in Oncology
Bringing Value from Big Data Analytics into Clinical Practice

Charles Mayo, Ph.D.
University of Michigan

Disclosure
Work is supported in part by a grant from Varian Medical Systems

It takes a village to make a Big Data Analytics Resource System

Randy Ten Haken, PhD
Marc Sosale, PhD
Dan McShan, PhD
Issam El Naqa, PhD
Jean Moran, PhD
Martha Matei, PhD
Scott Hadley, PhD
James Baller, PhD
Yue Cao, PhD
Dale Litzenberg, PhD

Awe Esbruch, MD
Jim Hayman, MD
Shivji Jhala, MD
Dawn Owen, MD
Reehampa Jagsi, MD
Michelle Marrawa, MD
Ted Lawrence, MD, PhD

Grant Wayburn
Carlos Anderson, PhD
Xiaping Chen
John Yap, PhD
Lynn Hockenbrock

Sue Merkel
Sherry Machnak
Denise Goodman
Sherry Machnak
Denise Goodman

Grace Sun
Alex Hayes
Parth Janni
**Data – Value Cycle**

**Manual**

- Extract, Transform, Load
- Minimal upstream coordination required
- Limited to relatively few numbers of patients (10s-100s)

**Automated Electronic**

- Extract, Transform, Load
- Large numbers of patients (>1000s)
- Lots of upstream coordination required:
  - access
  - standardization
  - people
  - resources
  - technical skills

---

**Better Patient Care**

- Multidisciplinary
  - Better Translational Research and PR
  - Incorporate data into routine practice

Build a strong foundation of data culture supporting Big Data initiatives for Radiation Oncology
So you want to build a Big Data Analytics Resource System?

Most of your effort is going to be in building and improving ETL processes:

- It takes a multi-disciplinary community that wants to make it real
- Invest in people with diverse skill sets
- Need commitment from leadership

University of Michigan – Radiation Oncology Analytics Resource

Be proactive on the Ethics of Access

- Acknowledge that data are people and can do harm
- Recognize that privacy is more than a binary value
- Develop a code of conduct for your organization, research community, or industry
Culture Shift: Standardize entry of Diagnosis and Staging -> Volume, Value

Data in the Electronic Medical Record

- Huge **volume** of text data available
- M-ROAR access (i.e. **velocity**) is fast (seconds)
- Potentially really **valuable** source of information
- The problem is **variability** ...
  - the solution is standardization

EMR Access + standardization -> Volume, Value

- Automate harvesting regular data entered into notes in EMR
- Presentation standardizations improve harvest-ability
We can do even better...

Culture Shift: Build templates in EMR with standardized schema for key data elements

- Standardized schema designed to function with regular expressions
- Physician selection from drop down lists of standard values &gt; Fast, Easy, Accurate

Requests for data tend to fall into three categories

<table>
<thead>
<tr>
<th>Type of query</th>
<th>Typical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice Quality Improvement (PQI)</td>
<td>Evidenced based approach to improving clinical processes and patient care</td>
</tr>
<tr>
<td>Translational Research</td>
<td>Provide data sets needed for publications and grants</td>
</tr>
<tr>
<td>Administrative Support</td>
<td>Ease access to data needed by front office e.g. Certificate of Need, Regulatory Groups, Institutional evaluation</td>
</tr>
</tbody>
</table>
Look at practical examples of using this resource to provide value... and what we've learned along the way.

Tableau Dashboards providing end user self-service.

---

**SRS and SBRT Utilization Analytics**

**Value Categories**
- Administrative Support
- Practice Quality Improvement

How can I look at how our SRS/SBRT program is evolving?

---

**Analysis of Lab Values**

**Value Categories**
- Practice Quality Improvement
- Translational Research

How can I look at trends in labs for a patient or look at labs for a set of patients?

**EMR > MRGAR**
- Batch processing possible
- Reduce work
- Integrate with treatment data
Patient Cohort Identification

How can I find a list of patients treated in a particular way?

Value Categories
- Practice Quality Improvement
- Translational Research
- Administrative Support

Treatment Timeline and Imaging Analytics

How long does it take to treat our patients and what imaging do we use?

Value Categories
- Practice Quality Improvement
- Translational Research
- Administrative Support

Can we use our historical DVH data when we are examining new treatment plans?

- Statistical DVH Dashboard
- Disease Site DVH Metric Summaries
- Practical Statistical Metrics
  - Generalized Evaluation Metric
  - Weighted Experience Score
  - Difficulty Ranking Score
  - Experienced based priorities
Statistical DVH Dashboard – Plan Summary Panel

Application is run from treatment planning system.
Uses visual and statistical metrics to compare this plan to historical plans.

Support for Machine Learning

- ML algorithms are data hungry: models and validation
- Need realistic representations of clinical distributions move from 10% to 1000% of patients
- Foundation for resolving differences between models
Community Science

- Variability in data and processes undermines reaching the real potential of Big Data

- No one institution can be the solution for these issues

- Viable solutions from community of individuals solving issues in their institutions then collaborating on shared solutions

\[ y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \]

Together, we are the equation

The information we have

More Information (Big Data) → More Complexity
Summary

- Analytics from Big Data fit readily into Clinical practice supporting
  - Translational Research
  - Practice Quality Improvement
  - Administrative Support

- Effort needed to build a Data Culture
  - Clinical practice changes
  - Support for access, extraction and curation

- Community Science
  - Development and publication of common standards
  - Multi-institutional data sets