Database sharing model for research and decision support in radiation therapy

Todd McNutt, Scott Robertson, Sierra Cheng, Joseph Moore, Harry Quon, Joseph Herman, Michael Bowers, John Wong, Theodore DeWeese

Disclosure:
Funding from Elekta, Philips and Toshiba

Personalized care using database of prior patients

How to best to treat individual patient?

Prediction of complications for early intervention?

Predicted response and toxicity?

Lab Values

Toxocities

Clinical assessments

Pathology

Diagnosis

Performance Status

Comorbidities

Patient History

Survival

Quality of Life

Disease Status

Radiotherapy

Surgery

Chemotherapy

FOLLOW-UP

TREATMENT

CONSULT

Project components

- Integration of data collection with clinical workflow
  - “Big Data” requires meaningful data
- Database design, security and distributed web-access
- Tools for query, analysis, navigation and decision support
  - Sample Questions and Uses
    • Toxicity Trending
    • DVH vs Toxicity
    • Automatic Treatment Planning and Quality
    • Prophylactic PEG use
Treatment Timeline

- **Simulation**
  - Demographics
  - Imaging
  - Motion

- **Planning**
  - Treatment Planning
  - Image Guidance

- **On Treatment**
  - Toxicity
  - Quality of Life

- **End of Treatment**
  - Acute Toxicity
  - Quality of Life
  - Patient Status

- **Follow Up**
  - Late Toxicity
  - Quality of Life
  - Patient Status
  - Disease Response

---

Data Collection in Clinic

- **Clinical Assessment**
- **Quality of Life**
- **Disease Status**

---

Chart Review (~40 per hour)
Extract, Transform, Load

- SQL Query
- Lab, Toxicity, Assessments

- Scripts, Python, DICOM
- DVH, OVH, Shapes

Oncospace tables and schema

Use of SQL DB reduces search to an SQL query
Informaticist and the Clinician

- Where clinical knowledge and informatics science meet?
- What is real knowledge?
- What is NEW knowledge?

The Vs of Big Data
**Viability and Value**

- Predictive factors must be accessible for new patients
- Prediction must be clinically valuable and extend the knowledge of the clinician

---

**Decision support for...**

- **SAFETY** can be improved by alerting users when patient treatment information deviates from normal.
- **QUALITY** can be improved by predicting how well you can do for a patient and seeking to achieve it.
- **PERSONALIZATION** occurs when physicians and patients can review results of prior similar patients and make decisions based on the data specific to the patients needs.

---

**Toxicity trends during and after treatment – detect outliers**

<table>
<thead>
<tr>
<th>Condition</th>
<th>During Treatment</th>
<th>Follow up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dysphagia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swallowing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucositis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflammation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xerostomia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dry Mouth</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Toxicity grade (0-5)

- Worsens after Tx for many patients then improves long term
- Heals after Tx for most patients
- Tends to be permanent
Shape-dose relationship for auto-planning

- More efficient plan optimization (10 fold)
- Normal tissue doses reduced (5-10%)
- Clinically released for Pancreatic Cancer

Flavors of Data
Depends on timeline

- **FIXED** already happened
  - Use to stratify patients

- **VARIABLE** influence outcome measures
  - Adjustable to influence outcome?

- **POPULATION** predict outcome measures
  - Predict outcome with FIXED and VARIABLE based on POPULATION?

Treatment Timeline

At what time point do we have enough data to make decision based on future prediction?

Input Variables => Prediction?
DVH, Toxicities and Grade distributions

Dysphagia and Xerostomia

Plugging Into Oncospace
Scott Robertson PhD
Voice Change

Bad DVH!

• DVH assumes that every sub-region of an OAR has the same radiosensitivity and functional importance to the related toxicity
• DVH assumes that each OAR is uniquely responsible for the overall human function related to the toxicity

Parotid Data
Classification with correlated features: unreliability of feature ranking and solutions

\[ F(\mathbf{U}) = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{y}^j - \mathbf{y}^j_{\text{mean}})^2 \]

Simulation of 1, 10 and 20 variables with a correlation of 0.9 with variable 3

Correlation is not Causation

The Texas Sharpshooter Procedure

Acknowledgments

- **JHU-RO**
  - Shihan Wu PhD
  - Kim Evans MS
  - Robert Jacques PhD
  - Joseph Marcus PhD
  - Scott Robertson PhD
  - Wayung Yang MS MD
  - John Wong PhD
  - Theodore DeWeese MD
  - GE Team
    - Joseph Marcus MD
    - Amy Hacker-Plitz PA
  - HKN Team
    - Harry Quan MD
    - Giuseppe Sampaio MD
    - Heather Turner MD
    - Jeremy Richmond MD
    - Anna Keiss MD

- **JHU-CS**
  - Russ Taylor PhD
  - Mika Kajihara PhD
  - Patricia Simon PhD
  - Jonathan Katzman

- **JHU-Physics**
  - Alex Szalay PhD
  - Tomas’ Bodnar PhD

- **Philips PROS**
  - Karl Bodnars

- **Erasmus**
  - Steven Pots PhD