Characterization of a novel detector using ROC

Elizabeth McKenzie Boehnke
Cedars-Sinai Medical Center
UCLA Medical Center
AAPM Annual Meeting 2017
New Ideas  Investigation  Clinical Practice
Investigating a novel detector using ROC Analysis

- Show how ROC analysis used as scientific tool to study new detectors
- Example of work with large area parallel plate ion chamber
- Discussion of how this can be generalized to other research
Evaluating Clinical Tools

• ROC Curves
  – Looks at how well a tool classifies both positives and negatives
    • Often, evaluations focus on catching errors without evaluating the false alarm rate
  – Flexibly compare multiple tools regardless of their metric
    • Just need to link to a binary output
  – Can quantitatively define a threshold for clinical use
    • Rather than using a “rule-of-thumb” approach
What is the IQM (Integrated Quality Monitor)?

• Purpose:
  – to perform real-time verification of treatment delivery

• Large area ion chamber with angled plates
  – Spatially dependent gradient along direction of MLC travel
  – Nearly unique signal from different collimation/fluence
What is the IQM (Integrated Quality Monitor)?

- Slides into head of linac
- Gives integrated counts as a function of control point
- Compares these integrated counts to a baseline measurement
Using the IQM

- Computer in control room where therapists can monitor the IQM signal
- If IQM signal is more than tolerance value (in terms of ± percent difference from baseline), alarm will sound

Example screenshot of IQM Monitor software showing IQM counts as a function of control point
Some aspects of this device

- It is monitoring for a failure
  - Needs specific cutoff values
- It is checking for a (hopefully) rare situation
  - Needs to be evaluated in a way which is robust to skew in the data
  - Potentially more passing cases than failing cases
Dealing with Skew

• Rare errors

• ROC curves are insensitive to changes in class distribution
  – OK to have more passing cases than failing cases in your experiment
  – ROC is plot of true positive vs false positive
Assessing the Sensitivity and Specificity: Managing Alarm Fatigue

- Previous publications have investigated the IQM’s sensitivity to MLC positioning errors, wrong energy, and wrong number of MUs
- In practice, need something that shows both good sensitivity and good specificity
  - Too many false alarms lead to **Alarm Fatigue**
Assessing the Sensitivity and Specificity

• To assess sensitivity and specificity, used ROC
  – Response: percent difference from baseline of IQM integrated count per control point
  – Binary variable: does the delivery have an error or not?

• Need to know which plans truly have an error
  – Gold standard
  – Generate error plans
Assessing the Sensitivity and Specificity

• To create error plans, used method of Steers et al 2016
  – Randomly perturb all in-field MLC positions by a distribution of errors ranging from ±1mm to ±5mm
  – For example, for the ±1mm error plan, the distribution of MLC errors per field will range from +1mm to -1mm
Assessing the Sensitivity and Specificity

- Used 5 original patient plans (3 brains, 1 liver, 1 lung)
- Created 5 additional MLC error plans per patient plan
Assessing the Sensitivity and Specificity

- For the analysis we used maximum magnitude of the percent difference per control point group
- Due to distributed nature of MLC errors and relevance to clinic
Distribution of percent differences & ROCs for good and less good plans

All of the MLC Error Levels (1mm to 5mm) were significantly different from their respective No-Error plans
Distribution of percent differences & ROCs for good and less good plans

All ROC curves showed good sensitivity and False Positive Rate.
Area Under the Curve

Area Under the ROC Curve as the MLC Error Increases

- AUC: Area under the ROC curve
- Gives an overall score for the classifier’s sensitivity and specificity
  - Loses some information
Optimal Tolerances

• Recall how the IQM data is given to the user...

If the measurement’s percent difference from baseline is greater than this → Alarm
Optimal Tolerances

Can find optimal tolerances from ROC curves, however due to inter-plan variation, more data needed to determine them.

<table>
<thead>
<tr>
<th>MLC Error (mm)</th>
<th>Liver Plan Optimal Cutoff (% from baseline)</th>
<th>Brain Plan1 Optimal Cutoff (% from baseline)</th>
<th>Brain Plan2 Optimal Cutoff (% from baseline)</th>
<th>Brain Plan3 Optimal Cutoff (% from baseline)</th>
<th>Lung Plan Optimal Cutoff (% from baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.6</td>
<td>10.8</td>
<td>11.0</td>
<td>23.3</td>
<td>22.0</td>
</tr>
<tr>
<td>2</td>
<td>2.0</td>
<td>10.8</td>
<td>10.8</td>
<td>22.9</td>
<td>21.1</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>17.4</td>
<td>11.5</td>
<td>20.8</td>
<td>16.3</td>
</tr>
<tr>
<td>4</td>
<td>6.9</td>
<td>12.1</td>
<td>12.1</td>
<td>21.8</td>
<td>21.9</td>
</tr>
<tr>
<td>5</td>
<td>6.8</td>
<td>14.7</td>
<td>12.4</td>
<td>21.6</td>
<td>16.4</td>
</tr>
</tbody>
</table>

![Example ROC Curve](chart.png)
Conclusions of Study

• IQM was able to tell the difference between plans with and without MLC errors (sensitivity and specificity)
• More work is needed to understand IQM response to different plan types in order to pick Optimal Tolerance values
Final Remarks

• ROC curves can be used to evaluate the sensitivity and specificity of emerging tools before becoming clinical
• They can be used to guide decisions on picking a relevant threshold
• They are flexible (only need binary outcome)
Acknowledgements

• Stephen Kry
• Benedick Fraass
• John Demarco
• Ke Sheng
Thank you
Receiver Operator Characteristic Curves

Different abilities to classify plans: possibly related to total MU (need more data to confirm)