The Science of Quality Assurance
Indicators & Technique Analysis

Marco Carlone, PhD, P.Eng, MCCPM
Medical Physicist, Princess Margaret Cancer Centre
Assistant Professor, University of Toronto
What is research?

In the broadest sense of the word, the definition of research includes any gathering of data, information, and facts for the advancement of knowledge. Reading a factual book of any sort is a kind of research. Surfing the Internet or watching the news is also a type of research.

Science does not use this word in the same way, preferring to restrict it to certain narrowly defined areas. The word ‘review’ is more often used to describe the learning process which is one of the underlying tenets of the rigid structures defining scientific research.

Definition of Research - How is Research Defined?

- The systematic investigation into and study of materials and sources in order to establish facts and reach new conclusions.
- Careful study that is done to find and report new knowledge about something.
- The activity of gathering information about a subject.
Quality Improvement vs. Research

The distinction between QI and research is an important one. There is a spectrum, and it can be blurry sometimes, but there are some key points (with legal implications!).

**QI:**
1. Intent is to improve current practice. For internal use only.
2. By definition, the data is confidential.
3. Action is within existing standards of care.
4. Institutional Review Board (IRB) approval is not necessary.

**Research:**
1. Intended to create generalized knowledge.
2. Desire to publish or present.
3. Testing new methods.
4. Needs IRB approval!
   (Click here to go to the Duke IRB website for more information)
Quality Improvement vs. Research

Quality Improvements

• Intent is to *improve current practice*. For internal use only.
• Action is *within existing standards of care*.
  – The knowledge is the same, but we can apply it in a better manner.

Research

• Intended to create *generalized knowledge*.
• Desire to publish or present.
  – Data must be relevant *outside the institution*.
• Testing new methods.
  – *Perhaps new standards of care?*
What is Patient Specific QA?

• Fundamentally, this is a beam diagnosis test.
• The purpose is to identify radiation beams that are “different” than the planned radiation beam.
• Most methods used to evaluate the result are physics & clinically based:
  – Dose & distance to agreement
  – Plan objectives, DVH, etc
• Are the usual metrics generalizable?
Dosimetry vs. Imaging

• Patient specific QA has been viewed as a dosimetry problem.

• From dose point of view:
  – How different is the measured dose from the planned dose?
  – Detectors, spatial resolution, etc.

• From image analysis point of view:
  – Is the measured beam fluence abnormal?
From a statistics point of view

• Statistical tools exist to measure how different two (1-dimensional) distributions are.
• Extended to 2-dimensions, this problem becomes very difficult to solve explicitly.
• Measures of dose difference at a point and distance to agreement:
  – indicate conformance,
  – difficult to generalize
Decision Theory

• The ROC Decision model
  – Receiver Operator Curve (ROC) is a plot of performance of a binary classifier system.

• Graphical tool allows quantification of best cut-off point.

• Also offers insight into where gains in sensitivity and specificity can be obtained.
Generalizing IMRT QA results

- Evidence that IMRT/VMAT QA results don’t translate well center to center.
- Letourneau, McNiven & Jaffray (IJROBP 2013)
  - Multi-institutional evaluation of IMRT/VMAT QA.
  - Results depend on the performance of many variables.
  - Variables interact with each other in ways that can be non-intuitive.
- Different centers may report IMRT/VMAT QA results that are similar, but the performance of their systems may be different.
  - Leaf calibrations, beam models, etc.
• 34 patients (single phase prostate cases)
• Half delivered normally.
• Other half delivered with known MLC errors
  – 1 mm, 2 mm, 3 mm
• Assume that the unperturbed delivery was “true” fluence pattern.
• Depending on passing criterion, measure rates of TP, TN, FP, FN.
Setting the detector’s operating point

![Diagram showing ideal cut-off percentage vs. average standard deviation (mm)]

- Ideal cut-off (%) plotted against \(<\sigma>(\text{mm})\)
- Graph includes data for 2%/2mm and 3%/3mm scenarios
- Inset ROC plots for different standard deviation values
Setting the detector’s operating point
Setting the detector’s operating point

An unbiased point that maximizes both sensitivity and specificity
Setting the detector’s operating point

ROC theory coupled with a decision model can choose ideal cut-off points with other goals: i.e. Minimize liability Quality/cost of detector
AUC

Test can be much less sensitive & specific

Test is highly sensitive & specific

$<σ>$ (mm)
AUC

<\sigma> (mm)

Poor Detectability

Good Detectability

2%/2mm

3%/3mm
Generalizing the result so that it can be compared to other tests, i.e., Leaf accuracy test.
The problem with $\gamma$

- “Gamma” is not a physical measurement, nor does it have statistical meaning.
- It has local significance, but it is difficult to interpret in a broad, multi-institution sense.
- It’s interpretation will always be controversial.
Summary

• QA activities are meant to help a specific local problem.
• Local problems are often relevant with a ‘narrow’ set of constraining factors.
• For the knowledge gained from these activities to be helpful in a broad sense, efforts can be made to choose metrics that have a generalized context.
• In the example of patient specific QA, thinking of the problem from a ‘detectability’ point of view led to quantification of some parameter thresholds that may be helpful in more general problem solving.