Methodologies for Evaluation of Effects of CAD On Users

Nicholas Petrick
Center for Devices and Radiological Health, U.S. Food and Drug Administration
AAPM - Computer Aided Detection in Diagnostic Imaging (CAD) Subcommittee

Learning Objectives

- To review the most common reader performance assessment techniques and associated performance metrics
- To review the AAPM CAD subcommittee's current thinking on best practices for performing and analyzing CAD reader performance assessment studies
- To understand the importance and limitations of various reader performance assessment approaches and metrics

Outline

- Introduce reader performance testing
- Preliminary work before starting a study
- Various types of studies and study designs
- Performance metrics
- Statistical analysis
- Summary
Reader Performance Testing

- Denotes a test (study) designed to estimate the performance of physicians using CAD as part of the diagnostic process
- Evaluating performance of physicians when aided by CAD

*http://www.imagingcentersofamerica.com/services/interpretation.php

Why reader performance testing?

- Reader studies are more indicative of clinical performance compared with standalone testing

CAD marks obvious lesions

- Marking only large lung nodules may not substantially improve physician performance
CAD marks subtle lesions

- Marking subtle lung nodules may improve physician performance
- Physician must be able to identify CAD mark as a nodule

Reader Performance Testing

Preliminary Work

Before Starting a Study

- Define how CAD should be used
- Conduct pre-study statistical analysis
  - Determine reader and case sample sizes
- Identify readers and image case set
- Establish reference standard
- Define mark labeling rules
How will CAD be used?

- Reading paradigm impacts design of reader performance study

**CAD Reading Paradigms**
- First reader
  - Physician reviews only regions or findings marked by the CAD device
- Sequential reading
  - Physician first conducts a full interpretation without CAD (unaided read)
  - Then re-evaluates with CAD aid (aided read)
- Concurrent reading
  - Physician performs full interpretation in presence of CAD
  - CAD marks are available at any time

Pre-study Statistical Analysis

- Pilot study is critical
  - Estimating effect size and variance components
  - Sizing reader study
    - No. of readers and no. of cases
    - Identifying problems in study design & protocol

**Freeware study sizing tools**

Example: iMRMC Tool

- Estimates nos. of readers and cases
- iMRMC plans to include a database of reader studies for references
Identify Readers & Cases

- **Readers**
  - Identify and sample from distribution of readers
  - Physicians expected to utilize device

- **Cases**
  - Test case set should match target population
  - General population for intended use
  - Sub-population clearly identified in the study aims
  - Inclusion/exclusion criteria clearly defined and justified
  - Distribution of known co-variates listed and any differences justified
  - Disease stage, lesion type, etc.

Other Pre-study Definitions

- **Reference standard**
  - Patient level
    - Whether or not disease is present
  - Lesion level
    - Location and/or extent of the disease

- **Mark labeling**
  - Rules for declaring a physician's mark as TP or FP
    - Only necessary when localizing lesions is study endpoint

Reader Performance Testing

MRMC Study Designs
Generalizability

- **Generalizability**
  - Study results generalize to a wider population
  - Ideally, results should generalize to clinical populations of patients, physicians, imaging hardware & protocols, reading environments, etc.
- **Controlled study results generalize to specific but limited populations**

Multi-Reader Multi-Case (MRMC)

- A controlled reader study where a set of readers interpret a common set of patient images
  - Typically under competing reading conditions
    - Readers unaided vs. readers aided by CAD

MRMC Study Designs

- Prospective
- Retrospective
Prospective Reader Studies

- **Randomized Control Trial**
  - CAD performance measured as part of actual clinical practice
  - Field testing a CAD device

Advantages

- Estimates clinical utility of CAD to readers as device is used in actual practice
- Good generalizability

Issues

- Generally require large patient population
- Low prevalence of disease in most CAD application areas

Types of Prospective Reader Studies

- Cross-sectional comparison studies
- Historical-control studies
- Double reading comparison studies
Cross-sectional Comparisons

- Physician interprets case without CAD assistance, records findings, interprets case again with CAD
  - Only possible when CAD is designed for sequential reading
    - Variation: Each case read independently by two physicians, one reading with CAD, one without
- Issues
  - Without CAD read may not match clinical practice

Historical-control Studies

- Compare physicians interpretation with CAD to their without CAD readings in a different (usually prior) time interval
- Advantage
  - Can be implemented directly within routine clinical practice
- Issues
  - Changes over study period confound CAD effect
    - Differences in patients, readers, interpretation process, etc.

Historical-control Studies

- Selection of the performance metric*
  - Cancer detection rate
    - May not correctly measure impact
  - Introduction of CAD may impact cancer prevalence in population of interest
- Alternate endpoints
  - Change in cancer stage, nodal status, etc.

Comparisons with Double Reading

- **Double reading**
  - Cases are read separately by two physicians
  - Increases detection sensitivity
  - Lowers specificity

- **Compare physicians interpretation with CAD to double reading**
  - Determine if single reading with CAD is as effective as double reading

Retrospective Reader Studies

- **Cases are collected prior to image interpretation and are read offline by the readers**

- **Most common approach for CAD is an enriched reader study design**
  - Population of cases enriched with patient known to be diseased

Retrospective Reader Studies

- **Advantages**
  - Substantially reducing no. of cases required to achieve statistically significant results
  - Allows for more rigorous study controls

- **Issues**
  - Impacts behavior of readers compared with clinical practice
  - Know their decisions don't impact patient care
  - Reader may become cognizant of enrichment relative to clinical practice

- **Control for reader behavior**
  - Compare performance to control modality
  - Reading without CAD
Retrospective Reader Studies

- **Warren-Burhenne Study Design**
  - Two separate studies
  - Retrospective study of CAD sensitivity to detect abnormalities in clinical practice
  - Estimated relative reduction in false negative (FN) rate with CAD
  - Study of the work-up rate of readers with and without CAD
  - Difference in work-up rate attributed to use of CAD
  - Study results may not be statistically interpretable


---

**Preferred approach**

- Evaluate all primary endpoints within a single study with physicians actually using the CAD
- [Se, Recall Rate] within one study
- Use ROC analysis

---

Reader Performance Testing

MRMC Performance Metrics
MRMC Performance Metrics

- Same as those discussed in prior talk on Standalone CAD Assessment

- Performance Metrics
  - ROC, LROC, JAFROC

- Figures of Merit
  - Area under the curve (AUC)
  - Partial area under the curve
  - Operating points
  - [Se, Sp], [Se, PPV], etc.

ROC Curve

Sources of Variability

- Cases
- Reader skill
ROC Curve with Operating Point

- Source of variability
  - Cases
  - Reader skill
  - Reader mindset
    - Different operating threshold for different readers

ROC Curve with Operating Point

Operating Point & ROC FOMs

- Hologic DBT reader studies
  - 2 reader studies
    - Compared 2D interpretation with 2D+3D interpretation
      - Study 1
        - 48 cancer cases
        - 264 non-cancer cases
        - 12 readers
      - Study 2
        - 51 cancers (48 same as in Study 1)
        - 259 non-cancer cases (171 same as in Study 1)
        - 15 readers (no overlap in readers with Study 1)

*http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevice AdvisoryCommittee/RadiologicalDevicePanel/ucm226660.htm
Reader Performance Testing

MRMC Statistical Analysis

- Sources of variability
  - Readers and cases

- Purpose of analysis
  - Determine statistically significant effects
Patient-Based MRMC Analysis Tools

- Dorfman, Berbaum and Metz (DBM)\(^1\)
- ANOVA analysis of jackknife pseudovalues
- Obuchowski-Rockette (OR) method\(^2\)
- ANOVA with correlated error analysis
- Directly models accuracy of each reader
- One shot MRMC method\(^3\)
  - Non-parametric approach
  - Based on mechanistic MRMC variance
  - Sum estimable moments to determine total variance


Location-based MRMC Analysis Tools

- Jackknife AFROC (JAFROC)\(^1\)
  - Discussed in prior talk on Standalone Assessment
- Region-of-Interest (clustered) ROC analysis\(^2\)
  - Divide patient data into regions


Region-of-Interest (clustered) ROC Analysis

- Non-clustered
  - Physicians scores enter 4-view set
- Clustered analysis
  - Physicians scores Right & Left pairs separately

\(^1\)http://www.wjco.com/content/figures/1477-7819-5-124-4.jpg
MRMC Analysis Freeware

- DBM software
  - http://perception.radiology.uiowa.edu/
- OR software
- One-shot software
- JAFROC software
  - http://www.devchakraborty.com/

Summary

- Reader performance testing evaluates performance of physicians when aided by CAD
  - Most often compared to unaided reading
- Statistical analysis/sizing freeware is available
- Pilot studies are critical for both
  - Sizing trials
  - Identify reading protocol issues

Additional Resources

- FDA Guidance for Industry and FDA Staff
  - Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions
    - http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm187277.htm
  - Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions
    - http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm187249.htm
Acknowledgements

- The mention of commercial entities, or commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such entities or products by the Department of Health and Human Services.