Image quality evaluation in CT

K.J. Myers[1], A. Badano[1], W. Jung[2], N. Petrick[1], L. Popescu[1], J. Vaishnav[2], R. Zeng[1]

[1]Division of Imaging and Applied Mathematics
Office of Science and Engineering Laboratories

[2]Division of Radiological Devices
Office of In Vitro Diagnostic Device Evaluation

CDRH FDA

Challenges to CT image evaluation

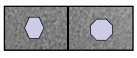
- CT systems are not linear and not shift-invariant
  - Resolution/sharpness depends on object size, contrast, location, and noise level
    - There isn't one MTF that describes the system transfer
  - Noise is colored
    - Pixel variance doesn't tell the whole story
  - Noise is nonstationary
    - Noise texture or NPS depends on location

All the more so for some iterative algorithms

Desired properties for performance evaluation method

- Task-based:
  - Performance of a specified task by a specified observer
- Possible tasks:
  - Detection of objects

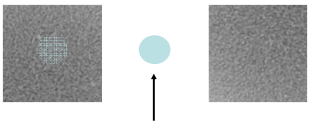
  - Discrimination of objects of different sizes or shapes



  - Estimation task: evaluation of the ability to measure a quantity such as tumor volume

Which image contains the signal?

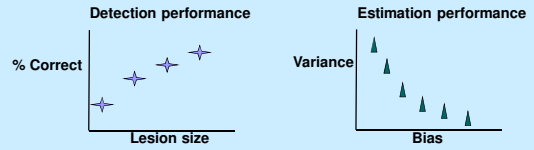
23% 1. The image on the LEFT
24% 2. The image on the RIGHT
25% 3. I can't see a thing from the back of the room.
27% 4. It's not visible in the front, either!



Signal example

Desired properties for performance evaluation method

- Task-based:
  - Performance of a specified task by a specified observer
- Possible tasks:
  - Detection of objects; discrimination of objects of different sizes; or even an estimation task (evaluation of the ability to measure a quantity)
- Objective: Figure of merit summarizes task performance



Desired properties for performance evaluation method

- Task-based:
  - Performance of a specified task by a specified observer
- Possible tasks:
  - Detection of objects; discrimination of objects of different sizes; or even an estimation task (evaluation of the ability to measure a quantity)
- Objective
  - Figure of merit is lesion detectability; or size discriminability; or estimation EMSE
- Reliable
  - Error bars are known and allow meaningful comparisons/conclusions
- Practical in terms of number of images, etc.
- Easily standardized

Assessment via human observers

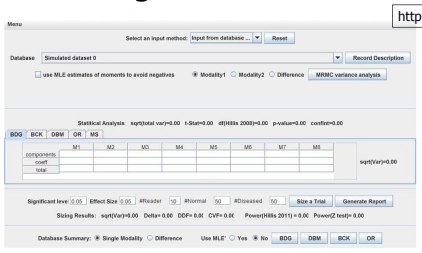
- Essential for clinical images and tasks
  - Expensive and time consuming
  - Observer variability can be considerable

Assessment via human observers

- Essential for clinical images and tasks
  - Expensive and time consuming
  - Observer variability can be considerable
- Requires random-effects or multivariate ROC
  - "MRMC" = Multi-Reader Multi-Case analysis
  - Gives total uncertainty in performance estimates from variability in images and observers (both skill and threshold)
  - Essential for testing significance of difference in competing modalities with multiple observers

iMRMC: Webpage and Software for Sizing an MRMC Clinical Trial

http://js.cx/~xin/index.html



- A resource for investigators designing a trial to compare two imaging modalities.
- Uses datasets from previous imaging trials to estimate power of new trial designs.
- Over time, database growth will benefit wide community of clinical trialists.

Special Review

Evaluating Imaging and Computer-aided Detection and Diagnosis Devices at the FDA

Brandon D. Gallas, PhD, Heang-Ping Chan, PhD, Carl J. D'Orsi, MD, Lori E. Dodd, PhD, Maryellen L. Giger, PhD, David Gur, ScD, Elizabeth A. Krupinski, PhD, Charles E. Metz, PhD, Kyle J. Myers, PhD, Nancy A. Obuchowski, PhD, Berkman Sahiner, PhD, Alicia Y. Toledano, ScD, Margarita L. Zuley, MD

Acad Radiol 2012; 19:463–477

How to evaluate the many system geometries or parameters of an iterative algorithm?

Assessment via model observers: surrogates or stand-ins for humans

- Significant literature validating models that predict human performance for simple detection/ discrimination tasks in images with variety of noise textures relevant to CT iterative reconstruction
- Enables system and algorithm evaluation without (or with fewer) human studies

Model observers for evaluation of image reconstruction algorithms

Since intent of algorithms is to display images to humans in useful form…

Models should be those that predict human performance ➔ Anthropomorphic models

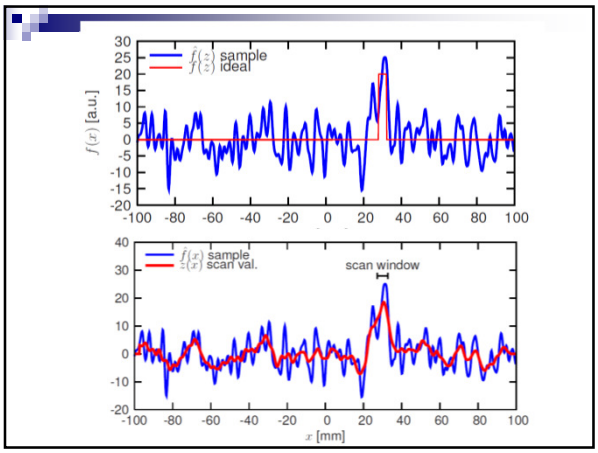
Channelized linear observers

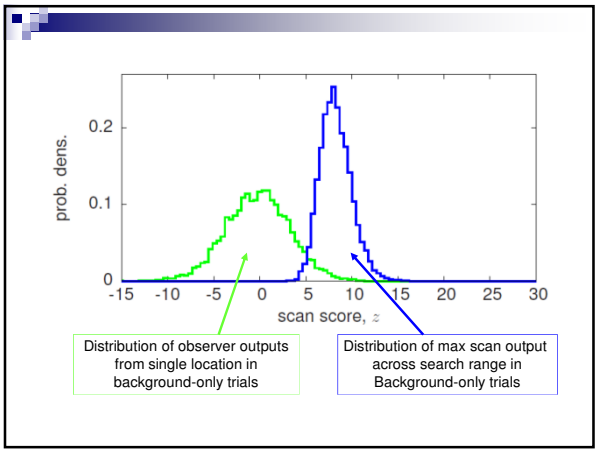
Channelized observer models

- 4-5 channels can give reasonable estimates of performance with 10-25 images
  - Will depend on # of signal realizations per image and their detectability
- May need to train model observer for each condition
  - Account for differences in image properties
- Software is available and more is coming
  - Arizona Image Quality Toolkit
  - CDRH Multiple Model Observer Calculator: http://code.google.com/p/mumoc/

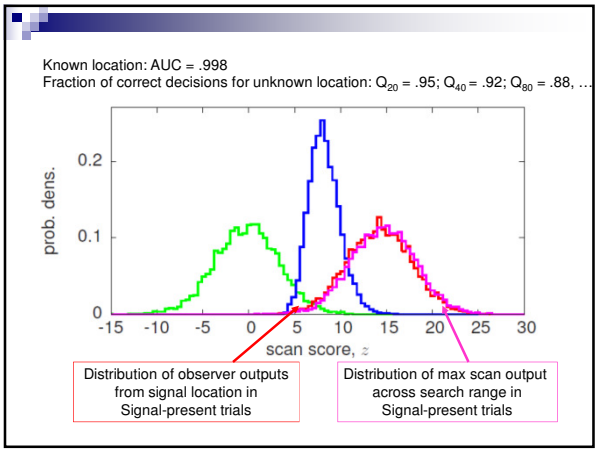
The case for a search task

- More like a clinical task
- Assesses how often the background/noise "looks like" the signal
- More generally, uncertainty in signal location (or size, shape) allows for more "dynamic range" in task SNRs available for given image set





Known location: AUC = .998
Fraction of correct decisions for unknown location: Q_{20} = .95; Q_{40} = .92; Q_{80} = .88, …



Why are standard phantoms not well suited to quantitative dose reduction studies?

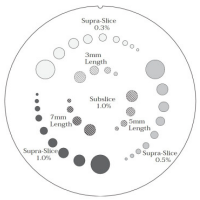
25% 1. Each signal size-contrast combination occurs once per image.

26% 2. The ability to define background-only ROIs is limited.

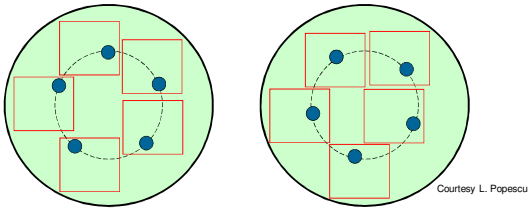
25% 3. An ROI containing just one of the signals would be small.

25% 4. All of the above.

Catphan®



Need for new phantom designs for OAIQ* studies



Courtesy L. Popescu

- Objects at same radial location with randomly placed ROIs ➔ search task
- No need for background-only ROIs for search task
- For 2AFC or ROC studies, obtain background ROIs from different locations in same slice, different slices in same scan, or different scans

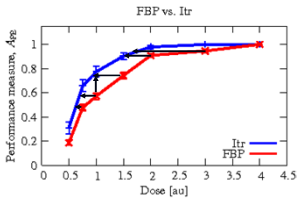
*Objective Assessment of Image Quality

Additional considerations

- Beware of non-relevant clues that lead to bias in IQ estimate
  - Avoid stitching artifacts, signal ROIs with different local background level or noise texture than background-only regions, etc.
- Means for avoiding these issues:
  - Rotate phantom between acquisitions
  - Randomly define ROIs to randomize signal location within them
  - Rotate/flip ROIs…

Quantifying dose reduction claims in CT

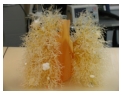
- Joint FDA-MITA task group developing framework for validation of claims:
  - Phantoms as stand-ins for the patient
  - Software for automated assessment
    - Rapid system evaluation without confounding factors of display and human inefficiency and variability
  - Statistical tools for measuring performance



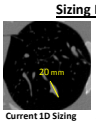
Next up: more interesting phantoms, tasks, etc.

- 3D tasks and observers
- Tasks related to temporal sampling (Fluoro, dynamic CT)
- Limited angle CBCT – tasks needed to assess limitations from artifacts
- Assessment of artifact reduction methds, e.g., metal implants
- Estimation tasks, e.g., tumor volume

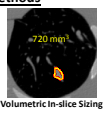
Anthropomorphic Thorax Phantom



Sizing Methods



Current 1D Sizing



3D Volumetric In-slice Sizing

Synthetic Nodules designed by CDRH

Spherical Ellipsoid Lobulated Spiculated

Thousands of CT images have been made publicly available by CDRH for use in software development & testing

Summary

- Joint FDA-industry collaboration on validation of dose-reduction claims for CT iterative reconstruction algorithms
  - Objective measures of image quality using phantoms and human or model observers
- Designing specific phantoms for OAIQ*
- Software (observers and performance metrics) for tasks with variability in signal location and other parameters

*Objective Assessment of Image Quality


Thank you