Quantitative Imaging Metrology: What Should Be Assessed and How?

Methods for Technical Performance Assessment

Nicholas Petrick
CDRH/OSEL/DIDSR
U.S. Food and Drug Administration

Outline of Talk

- Purpose
- Main components of technical assessment
- Example Analysis
- Summary

Purpose

- Discuss QIBA consensus on technical assessment of QIB performance*
  - Include example analysis from our research

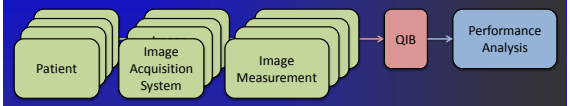
 *Raunig et al., Stat Methods Med Res, 2014

Why FDA?

- QIB are important growing area of interest
- CDER (drugs)
  - Qualification of QIBs for use in drug trials
- CDRH (devices)
  - Potential for specific QIB device claims
  - Develop least burdensome assessment methods

*Raunig et al., Stat Methods Med Res, 2014

Components of a QIB



- Factors influencing reliability of a QIB
  - Patient factors
    - Natural variation in lesions/patients
    - Variation among patients
  - Acquisition parameters
    - Slice thickness, collimation, exposure, reconstruction, etc.
    - Variation in imaging platform
  - Software effects
    - Automated/semi-automated
    - Variation in measurement tool

QIB Technical Assessment

- QIBs typically evaluated for reliability under study conditions
- Problem: QIB performance evaluation often different from study-to-study
  - Partly due to differences in data set, etc.
  - Also due to analysis differences
    - *Ad hoc* metrics
    - Inconsistent use of statistical metrics
    - Inconsistent terminology

QIBA Technical Assessment Group*

- Provide common framework for researchers to assess technical performance
- Ensure reliability of image measurement
  - Subject QIBs to same rigor as other quantitative clinical measurements

*Raunig et al., Stat Methods Med Res, 2014

Designing QIB analysis plan*

Step 1: Define QIB and its relationship to quantity to be measured (measurand)
Step 2: Define question to be addressed
  - Hypothesis or bounds on performance

Step 3: Define the experimental unit
  - Lesion-level, patient-level

Step 4: Define statistical measures of performance
  - What are the metrics for bias, repeatability, reproducibility?

Step 5: Specify elements of statistical design
  Reference, reproducibility conditions, etc.

Step 6: Determine the data requirements
  - Patient population, type of images, sample size, etc.

Step 7: Collect data and perform statistical analysis

*Raunig et al., Stat Methods Med Res, 2014

Basic Technical Assessment Components

- Bias/Linearity
- Repeatability
- Reproducibility

QIB Technical Assessment

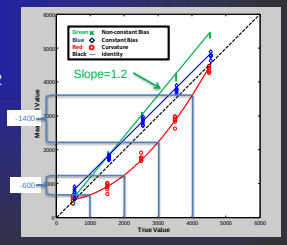
Bias and Linearity

Bias and Linearity

- Relationship of QIB to reference standard (measurand)
  - $\mathrm{E}(\hat{X}|X = x) = f(x)$
    - $\hat{X}$: Estimate
    - X: Reference standard (measurand)
- Linearity
  - On average, change in x reflects a proportional change in $\hat{X}$

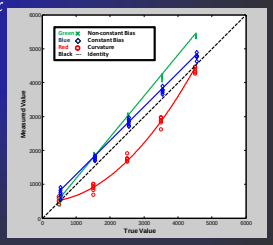
Bias and Linearity

- Linearity
  - Proportional change
    - True (green) : 1.0
    - Measured (green): 1.2
- Nonlinear
  - True (red) 1.0
  - Measured
    - 1000: 0.6
    - 3000: 1.4



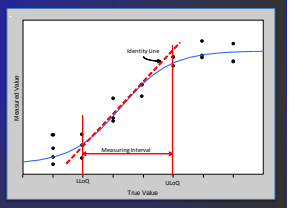
Bias

- $Bias(\hat{X}) = E(\hat{X}|X = x) - x$
- Types
  - Unbiased
    - $Bias(\hat{X}) = 0$
      - Slope=1
      - Intercept=0
  - Constant bias
    - $Bias(\hat{X}) = B$
      - Slope=1
  - Non-constant bias
    - $Bias(\hat{X}) = B(x)$
      - Depended on truth



Bias and Linearity Assessment

- Visually assess measured vs reference
  - Define limits of quantitation
- Bias analysis
  - Utilize multiple replicates for multiple "truth" values
- Linearity analysis
  - Test for linearity
    - Sequential polynomial testing



QIB Technical Assessment

Repeatability

Repeatability

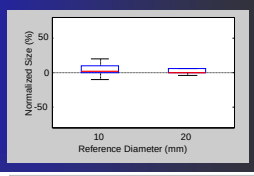
- Test-retest within-subject variability
  - Within-subject variability
    - Variability across repeat observations nested within patients

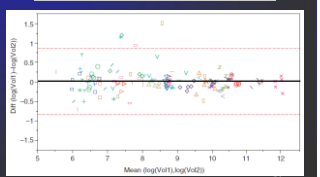
Repeatability Metrics

- Within-subject variance
  - $\sigma_w^2 = E_p\left(E_r\left((y_{ij} - \mu_i)^2\right)\right)$
    - Repeat observations, $j = 1, \dots, k$
    - Patient, $i = 1, \dots, n$
- Repeatability Coefficient
  - Variance of difference in two independent measurements
    - $RC = 1.96\sqrt{2\sigma_w^2}$, for normal data
  - Limits of Agreement
    - $LOA = [-RC, RC]$
- Others
  - Intra-class correlation (ICC)
  - Within-subject coefficient of variation (wCV)

Repeatability Plots

- Reference known
  - Box-whisker plot
- Reference unknown
  - Bland-Altman plot
    - Plot of difference against mean of two repeated measurements





QIB Technical Assessment

Reproducibility

Reproducibility

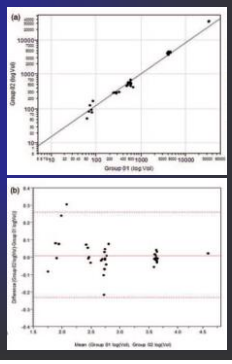
- Variability of QIB under differing experimental conditions
- Evaluate QIB performance under varied but controlled conditions
  - Potential reproducibility conditions
    - Across scanners
    - Across QIB measurement tools
    - Across sites
    - Across readers/clinicians

Reproducibility Metrics

- Within-subject reproducibility variance
  - $\sigma^2_{Reproducibility} = \sigma^2_{Repeatability} + \sigma^2_{between-conditions}$
- Reproducibility Coefficient
  - Variance for difference in two independent measurements
    - $RDC = 1.96\sqrt{2\sigma^2_{Reproducibility}}$, for normal data
  - Limits of Agreement
    - $LOA_{RDC} = [-RDC, RDC]$
- Others
  - Concordance correlation (CCC)

Reproducibility Plots

- 2 Factors
  - Scatter plots
    - Regression
      - Slope= 0.97
      - Intercept= 0.09
  - Bland-Altman plot
    - RDC= 0.234
- ≥2 Factors
  - Box-whisker plot
  - …



QIB Technical Assessment

Phantom Study Example

Purpose

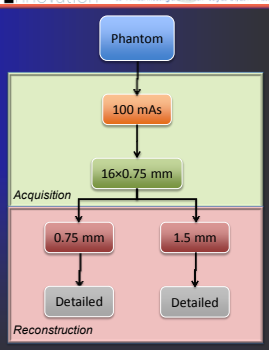
- Evaluate technical performance of a nodule volume estimation tool
  - Segmentation-based nodule volume estimation tool
  - Reproducibility condition
    - CT slice thickness ranges includes 075 mm & 1.5 mm

Study Design

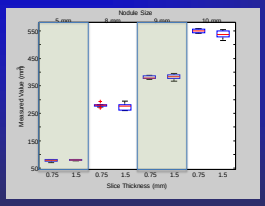
- Phantom
  - Thorax phantom with vascular insert
- Synthetic nodules
  - 4 spherical nodules
    - 5, 8, 9, 10 mm
    - +100 HU

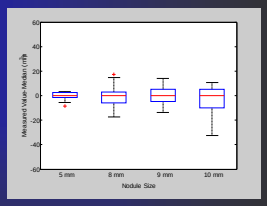
Study Design

- Image collection protocols
  - 10 repeat acquisitions



Bias/Linearity



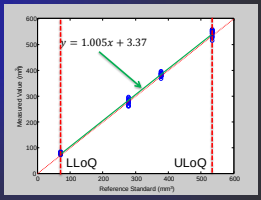


- Visually assess data
  - Evaluate means/variances
    - Is data transformation necessary/appropriate?

QIB Technical Assessment

Bias/Linearity

- Visually assess measured vs reference
  - Linear Regression
    - $y = 1.005x + 3.37$
  - $Bias_{Est} = 4.79\ mm^3$



7/21/2014

QIB Technical Assessment

Results

- Repeatability/Reproducibility

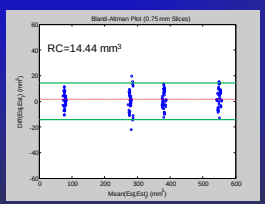
	Slice Thickness	
	0.75 mm	1.5 mm
RC	14.4 mm^3	27.6 mm^3
RDC	29.0 mm^3	

  - Want to calculate CI on values

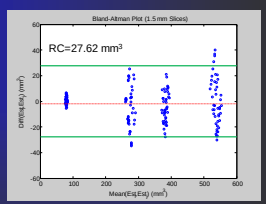
7/21/2014 42

QIB Technical Assessment

Repeatability Analysis



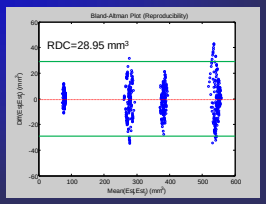
0.75 mm Slices



1.5 mm Slices

7/21/2014 43

Reproducibility Analysis



Summary

- Main components of QIB technical assessment
  - Bias/linearity analysis
  - Repeatability analysis
  - Reproducibility analysis
  - Others
    - Identification of significant factors/subgroups
    - ….
- Challenging to maintain consistency across studies
  - Phantom/clinical data
  - Transformation of data
  - Reference standard
  - Is test-retest data available?
  - Reproducibility conditions
  - ….

Acknowledgement

- Qin Li and Marios A. Gavrielides for their efforts on the FDA QIB projects
- Qin Li for her efforts in developing this presentation