Memorial Sloan Kettering Cancer Center

Assessing New Quantitative Imaging Biomarkers

Chaya Moskowitz, PhD
Department of Epidemiology & Biostatistics

With thanks to Todd Alonzo

Financial disclosures: None

1

Outline

- Steps to evaluate whether an imaging biomarker is useful
- Evaluating technical performance with an emphasis on precision

Memorial Sloan Kettering Cancer Center

2

Quantitative Imaging Biomarkers

- Feature or characteristic that is objectively measured from a medical image and is an indicator of a normal biological process, pathogenic process, or response to therapeutic intervention.
- Continuous scale

Sullivan et al. *Radiology* 2015

Memorial Sloan Kettering Cancer Center

3

Useful Quantitative Imaging Biomarkers

Must have:

- Analytic validity
  - Technical performance; Does the imaging biomarker measure what it is supposed to measure?
- Clinical validity
  - Is the imaging biomarker associated with the clinical (patient) outcome?
- Clinical usefulness
  - Does the imaging biomarker have a positive impact on patients or public health?

Memorial Sloan Kettering Cancer Center

4

Analytic Validity

- Early-phase studies
  - Preclinical, laboratory studies
  - Early clinical development

Memorial Sloan Kettering Cancer Center

5

Analytic Validity

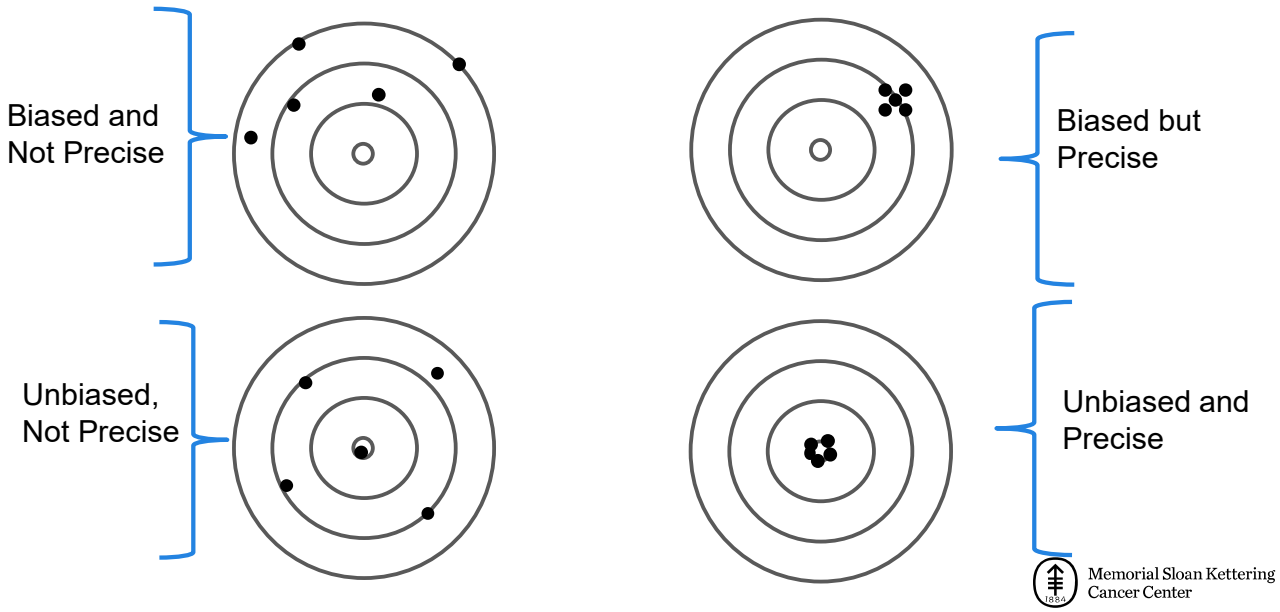
- Early-phase studies
  - Preclinical, laboratory studies
  - Early clinical development

- Study endpoints and metrics
  - Bias, analytic accuracy
    - Mean differences between measurement and truth
    - Analytic sensitivity, specificity, ROC curves
  - Precision
    - Repeatability, reproducibility

Memorial Sloan Kettering Cancer Center

6

Precision vs. Bias



Memorial Sloan Kettering Cancer Center

7

Importance

- Potential utility of an imaging biomarker can be greatly impacted by lack of precision
- Poor precision can make measured change in biomarker difficult to interpret
- Developing precise quantitative imaging biomarkers can be difficult
- Acceptable magnitude depends on use
  - High precision should be a necessary component of any procedure intended for diagnostic use

Memorial Sloan Kettering Cancer Center

8

Sources of Bias and Variability

- Patient-related
  - Disease or treatment-related
  - Other biophysiological sources
- Imaging system-related
  - Scanner-related
  - Human element

Memorial Sloan Kettering Cancer Center

9

Repeatability and Reproducibility

- **Repeatability**: consistency of results when same imaging biomarker is assessed at short intervals on same subjects using same equipment, same reader, in same center
- **Reproducibility**: consistency of results when same imaging biomarker is assessed at short intervals on same subjects using different equipment, different reader, or in different centers

Barnhart and Barboriak, *Translational Oncology* (2009)

Memorial Sloan Kettering Cancer Center

10

Study Designs

- **Repeatability**:
  K repeated measurements ($K \geq 2$) on *n* subjects
  - Identical conditions
  - Test-retest, "coffee-break studies"
  - *n* should be at least 35*
- **Reproducibility**:
  K methods/readers measure ($K \geq 2$) *n* subjects
  - Vary component(s) systematically
  - Method comparison
  - *n* should be at least 35

*Obuchoswki and Bullen, 2018

Memorial Sloan Kettering Cancer Center

11

Metrics for Assessing Precision

- Descriptive statistics
  - Means, variances, correlations
- Plots
  - Pairwise scatter plots
  - Bland-Altman plots (Bland and Altman, *Lancet* (1986))
    - Plot of difference vs average
    - Mean difference
    - 95% Limits of Agreement (mean difference ± 2 *x* standard deviation)
- Primary metrics usually rely on:
  - Absolute differences between measurements
  - Components of variance

Memorial Sloan Kettering Cancer Center

12

Assessing Repeatability

- Frequently based on within-subject standard deviation, σ_W
  - Repeatability coefficient : RC = 2.77 σ_W
    - Used as a cutoff for distinguishing real change from measurement noise
    - Value under which absolute differences between repeated measurements on same patient should fall with 95% probability
  - Within-patient coefficient of variance: wCV= σ_W / μ
    - Variation in repeated measurements relative to typical meaurements

Memorial Sloan Kettering Cancer Center

13

Example of Repeatability Study

ACRIN American College of Radiology Imaging Network

Quantitative Imaging Biomarkers Alliance RSNA

AMERICAN COLLEGE OF RADIOLOGY IMAGING NETWORK and QUANTITATIVE IMAGING BIOMARKERS ALLIANCE

ACRIN 6701

Repeatability Assessment of Quantitative DCE-MRI and DWI: A Multicenter Study of Functional Imaging Standardization in the Prostate

Primary Aim: Determine the test-retest performance, assessed by the repeatability coefficients (RC) of K^{trans} and $IAUGC90^{bn}$ and measured by median pixel values of the whole prostate.

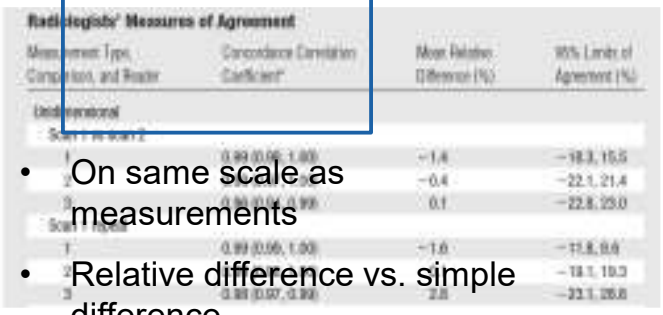
Memorial Sloan Kettering Cancer Center

14

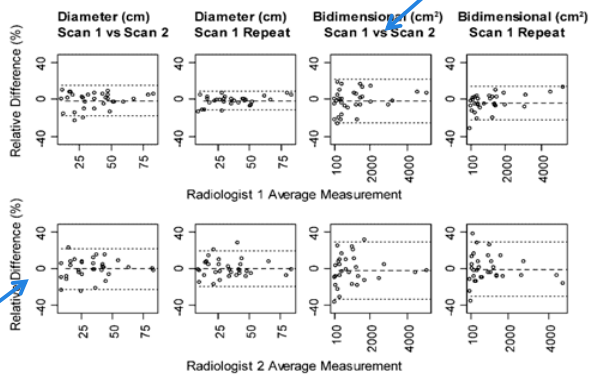
Example of Repeatability Study

Radiology

Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non–Small Cell Lung Cancer[1]



Radiologists' Measures of Agreement



- On same scale as measurements
- Relative difference vs. simple difference
- You interpret!

Zhao *et al.*, *Radiology* (2009)

Memorial Sloan Kettering Cancer Center

15

Assessing Reproducibility

- Frequently based on between-subject standard deviation, σ_B
  - Intraclass correlation coefficient
    - ICC = $\sigma^2_B / (\sigma^2_B + \sigma^2_W)$
    - Interpretation: Proportion of total variance due to the different readers/methods
  - Concordance correlation coefficient (Lin, Biometrics (1989))
    - $\rho_c = (2\sigma_{X_1X_2}) / (\sigma^2_{X_1} + \sigma^2_{X_2} + (\mu_{X_1} - \mu_{X_2})^2)$
    - Interpretation: Quantifies agreement between two measurements

Memorial Sloan Kettering Cancer Center

16

Example of Reproducibility Study

Reproducibility of Measurement of Apparent Diffusion Coefficients of Malignant Hepatic Tumors: Effect of DWI Techniques and Calculation Methods

Interobserver Agreement for ADC Measurement Presenting With ICC

		Breath-hold DWI		Respiratory-triggered DWI					
				Two b-value method			Multiple b-value method		
		ICC	LOA[a]		ICC	LOA[a]		ICC	LOA[a]
First	$ADC_{0/500}$	0.979 (0.921-0.993)	11.3	$ADC_{0/500}$	0.918 (0.803-0.967)	14.3	ADC_{0-500}	0.953 (0.884-0.981)	12.1
Second		0.974 (0.934-0.990)	11.8		0.925 (0.736-0.974)	15.9		0.917 (0.760-0.969)	16.4
First	$ADC_{50/500}$	0.983 (0.942-0.994)	11.4	$ADC_{50/500}$	0.969 (0.922-0.988)	12.6	ADC_{50-500}	0.972 (0.928-0.989)	12.1
Second		0.964(0.911-0.986)	12.9		0.878 (0.697-0.992)	21.2		0.889 (0.772-0.957)	20.2
First				$ADC_{0/1000}$	0.974 (0.934-0.990)	8.2	ADC_{0-1000}	0.979 (0.947-0.992)	7.7
Second					0.803 (0.555-0.919)	20.7		0.803 (0.555-0.919)	23.2

Numbers in parentheses are 95% confidence interval.

Kim *et al.*, *J of Magentic Resonance Imaging* (2012)

Memorial Sloan Kettering Cancer Center

17

Other Considerations

- Many other possible methods
- Estimation rather than testing
  - P-value less interesting
  - Confidence intervals

Memorial Sloan Kettering Cancer Center

18

Conclusions

- Critical to assess analytic validity; many studies do not rigorously assess analytic validity
  - Consistency of results when imaging biomarker assessed at short intervals on same subjects
  - Primarily early-phase studies, but methods may be useful for later-phase studies as well
  - Design studies to evaluate both repeatability and reproducibility
- Equally critical to assess both clinical validity and clinical usefulness of an imaging biomarker

Memorial Sloan Kettering Cancer Center

19