Transitioning QIBs into Clinical Trials:
Practical Implementation Requirements

Nancy A Obuchowski, PhD
Cleveland Clinic Foundation

Acknowledgement: P. David Mozley, Dawn Matthews, Andrew Buckler, Jennifer Bullen, and Edward Jackson

Questions:

- How can our knowledge about measurement error be used to better design clinical trials?
- What do we really need to know about the QIB?

Test-Retest Studies:
Estimate repeatability

Phantom Studies:
Estimate bias, and assess linearity

Reproducibility Studies:
Estimate effect of imaging methods on precision and bias

Level 1: Technical Performance Validation

Test-Retest Studies:
- Estimate repeatability

Phantom Studies:
-Estimate bias, assess linearity

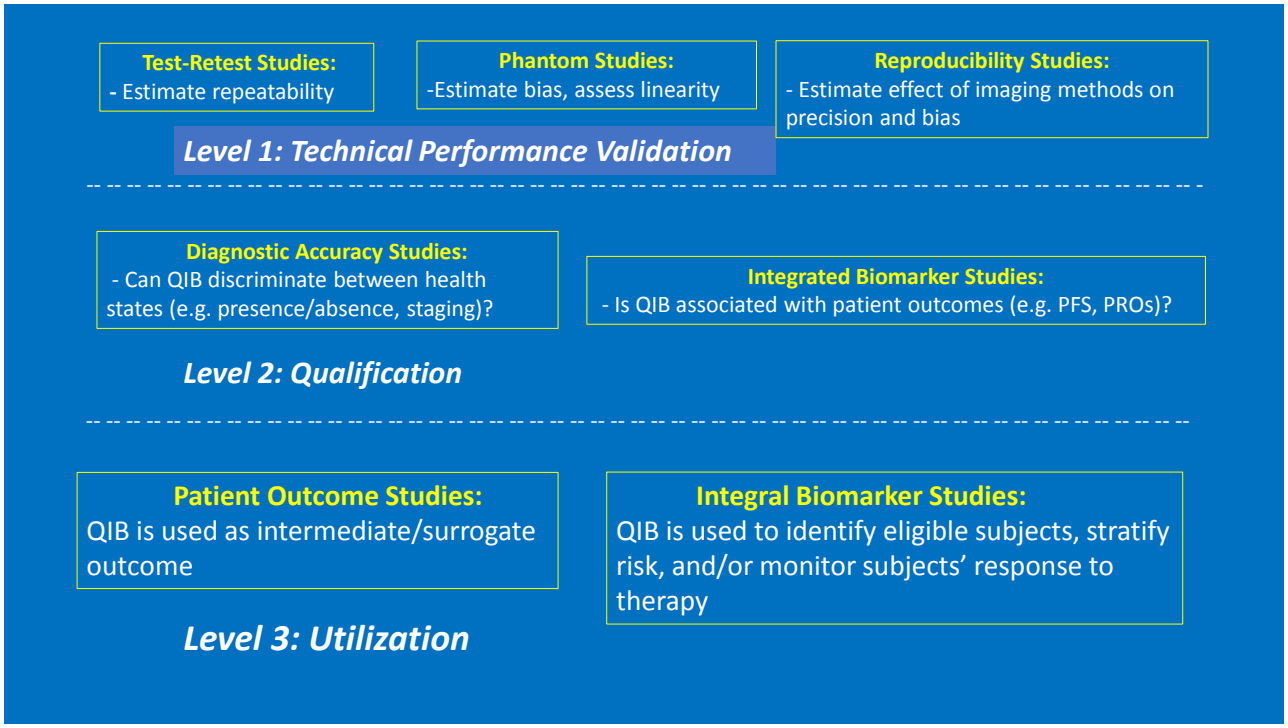
Reproducibility Studies:
- Estimate effect of imaging methods on precision and bias

Level 1: Technical Performance Validation

Diagnostic Accuracy Studies:
Can QIB discriminate between health states (e.g. presence/ absence of disease, staging)?

Integrated Biomarker Studies:
Is QIB associated with patient outcomes (e.g. Progression-free survival (PFS), Patient-reported outcomes (PROs))?

Level 2: Qualification



Example 1 (Qualification): Can ultrasound elastography discriminate subjects with liver cirrhosis (stage F4) from those without cirrhosis?

Is shear wave speed a diagnostic biomarker?

Example 2 (Qualification): Does the change in CT lung nodule volume after two weeks of treatment predict patient outcome?

Is the change in CT tumor volume a potential monitoring biomarker?

Example 3 (Utilization): Consider a RCT of Alzheimer's patients, comparing accumulation of amyloid over two years in subjects undergoing a neuroprotective treatment vs. subjects being treated symptomatically.

SUVr is used as study endpoint.

Example 4 (Utilization): SPECT specific binding ratio (SBR) in the posterior putamen is used as an eligibility criterion for identifying Parkinson's disease subjects likely to benefit from a new intervention.

SBR is used as a selection biomarker.

How does QIB measurement error affect clinical trials?

1. Attenuates ability of biomarker to discriminate between health states or predict outcome
   - Less power (not able to qualify the QIB)
2. Leads to incorrect decisions
   - Misinterpret measurements or change in measurements (not able to utilize the QIB)

How does QIB measurement error affect clinical trials?

1. Attenuates ability of biomarker to discriminate between health states or predict outcome
   - Less power (not able to qualify the QIB) → re-calculate sample size, accounting for QIB's measurement error
2. Leads to incorrect decisions
   - Misinterpret measurements or change in measurements (not able to utilize the QIB) → use the QIB measurement to construct 95% CI for true value

Correction to Sample Size:

$$\#\text{ patients needed} = N_X(\hat{\beta}_1^2\hat{\sigma}_b^2 + \hat{\sigma}_\epsilon^2)/\hat{\beta}_1^2\hat{\sigma}_b^2$$

↑ sample size if there was no measurement error

Obuchowski et al, JNCI *in press*

Correction to Sample Size:

patients needed $= N_X(\hat{\beta}_1^2\hat{\sigma}_b^2 + \hat{\sigma}_\epsilon^2)/\hat{\beta}_1^2\hat{\sigma}_b^2$

regression slope of measurements on true values

(we often assume slope=1 but critical that we test that)

Obuchowski et al, JNCI *in press*

Correction to Sample Size:

patients needed $= N_X(\hat{\beta}_1^2\hat{\sigma}_b^2 + \hat{\sigma}_\epsilon^2)/\hat{\beta}_1^2\hat{\sigma}_b^2$

between-subject variance

Obuchowski et al, JNCI *in press*

Correction to Sample Size:

$$\text{\# patients needed} = N_X(\hat{\beta}_1^2\hat{\sigma}_b^2 + \hat{\sigma}_\epsilon^2)/\hat{\beta}_1^2\hat{\sigma}_b^2$$

within-subject variance
(from test-retest studies)

Obuchowski et al, JNCI *in press*

95% CI for the true value (to account for measurement error):

$$(Y_i - \hat{\beta}_0)/\hat{\beta}_1 \pm 1.96 \times \hat{\sigma}_\epsilon / \hat{\beta}_1$$

fixed (mean) bias

Obuchowski et al, JNCI *in press*

95% CI (to account for measurement error):

$$(Y_i - \hat{\beta}_0)/\hat{\beta}_1 \pm 1.96 \times \hat{\sigma}_\epsilon / \hat{\beta}_1$$

regression slope of measurements on true values

Obuchowski et al, JNCI *in press*

95% CI (to account for measurement error):

$$(Y_i - \hat{\beta}_0)/\hat{\beta}_1 \pm 1.96 \times \hat{\sigma}_\epsilon / \hat{\beta}_1$$

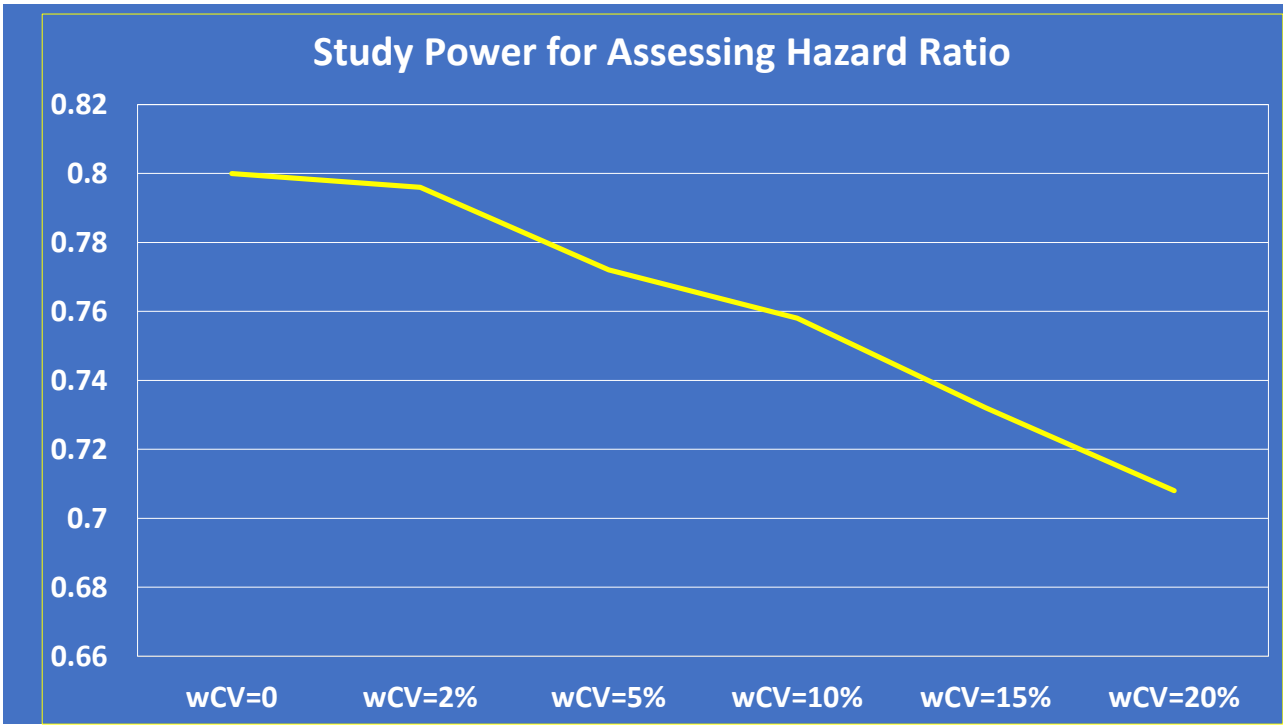
test-retest SD

Obuchowski et al, JNCI *in press*

Qualification Example: Consider a clinical trial of a new therapeutic intervention for lung cancer.

Primary objective: Compare progression free survival (PFS) of subjects in new vs. standard trt

Secondary objective: Test if change in CT lung nodule volume after two weeks predicts PFS in smaller subpopulation.



QIBA profile for CT tumor volume change in advanced disease:

wCV~8.5% for tumors with longest diameter 50-100mm

- *this allows different scanners and readers at two time points*

QIBA profile for CT tumor volume change in advanced disease:

Ignoring measurement error, N=54 subjects would be accrued to construct 95% CI of width ±0.3 for hazard ratio.

Accounting for wCV=8.5% and slope=1, we need to recruit N=62.

Trade-off between level of standardization and study practicality

	Quantitative Assessment				Qualitative
	Same scanner, Same reader, Same analysis tool	Same scanner, Different reader, Same analysis tool	Different scanner, Different reader, Same analysis tool	Different scanner, Different reader, Different analysis tool	RECIST
wCV	**2.9%**	**3.6%**	**8.5%**	**15.5%**	**---**
Study Power*	**~80%**	**79%**	**77%**	**71%**	**62%**

* A sample size of N=54 is needed when no measurement error is present for 80% power, 5% type I error to detect a HR>1.

Obuchowski et al, JNCI *in press*

Utilization Example: SPECT specific binding ratio (SBR) in the posterior putamen used as eligibility criterion for Parkinson's trial.

SBR≤1.2 → eligible for study (likely to benefit from new trt)

SBR >1.2 → excluded from study (unlikely to benefit)

QIBA's Profile on quantifying dopamine transporters with ^{123}Iodine-labeled ioflupane

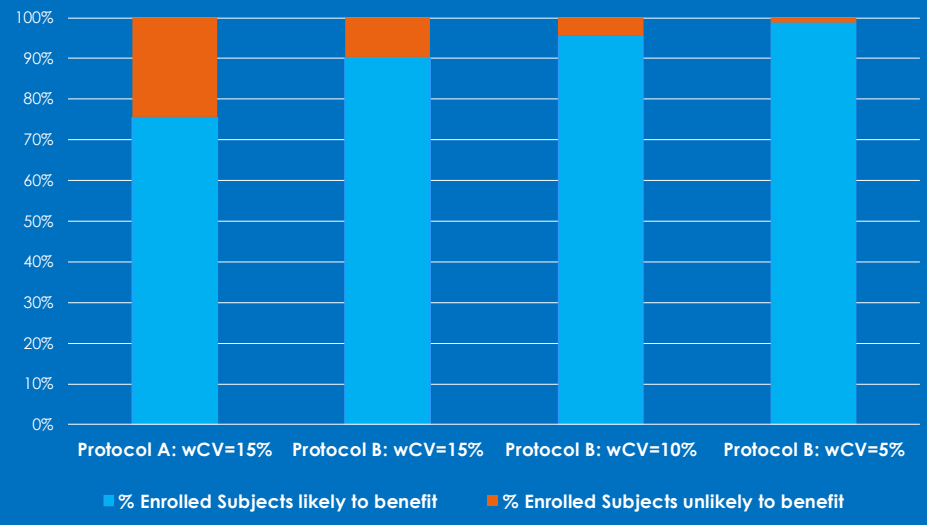
1. SPECT SBR measurements have wCV=15%
2. Regression slope ~ 1.0

Furthermore, assume negligible fixed (mean) bias.

Protocol A: Ignore measurement error and enroll patients if SBR $\leq$1.2

Protocol B: Use our knowledge of the technical performance of SBR measurements to construct 95% CI for true value. Enroll patients if CI for true value suggests they are eligible

% Enrolled Subjects Likely To Benefit From Intervention



Conclusions: ***Implementing QIBs in Clinical Trials***

1. Ignoring measurement error leads to:
   - low power in clinical trials
   - misinterpretation of QIB measurement

Conclusions: ***Implementing QIBs in Clinical Trials***

1. Ignoring measurement error leads to:
   - low power in clinical trials
   - misinterpretation of QIB measurement

2. If you know the measurement error:
   - recalculate sample size
   - use 95% CI for true value instead of QIB measurement
   - assess trade-off between standardization and practicality

Implementing QIBs in Clinical Trials

For *Qualification* studies, we need to know:
- test-retest variance (wSD, wCV)
- linearity exists and magnitude of slope

- reproducibility (to assess trade-off in standardization/practicality)

For *Utilization* studies, we also need to know:
- mean (fixed) bias