Assay validation and reproducibility considerations for biomarkers used in drug development

Evidentiary Considerations for Integration of Biomarkers in Drug Development: A Workshop co-sponsored by FDA and M-CERSI

Baltimore, MD

Lisa M McShane, PhD
Biometric Research Branch
Division of Cancer Treatment and Diagnosis

NIH NATIONAL CANCER INSTITUTE

August 21, 2015

Disclosures

I have no financial relationships to disclose.

- and -

I will not discuss off label use and/or investigational use in my presentation.

- and -

The views expressed represent my own and do not necessarily represent the views or policies of the National Cancer Institute.

Outline

1. Uses of biomarkers in drug development
2. Assay fit-for-purpose
3. Description of the assay
4. Assay performance assessment
   - Specimens & pre-analytic factors
   - Analytic performance

Uses of biomarkers in drug development

Uses of biomarkers in drug development

Biomarker Use	Drug Development Objective
Target validation	Demonstrate that a potential drug target plays a key role in the disease process
Early compound screening	Identify compounds with the most promise for efficacy and safety
Pharmacodynamic assays	Determine drug activity; select dose and schedule
Patient selection	In clinical trials, patient selection (inclusion/exclusion)
Surrogate endpoint	Use of an alternative outcome measure which can be measured sooner, less invasively, or with less inconvenience or cost, in place of the long-term primary endpoint to determine more quickly whether the treatment is efficacious and safe in drug regulatory approval

Adapted from IOM Biomarkers & Surrogates Report, 2010

Increasing demands on a biomarker assay

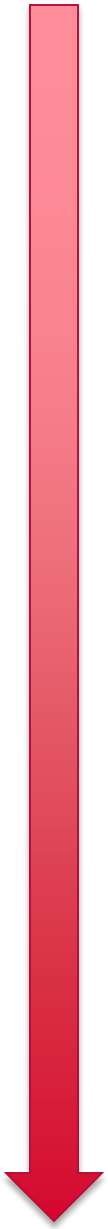
Target validation

Early compound screening

Pharmacodynamic assays

Patient selection

Surrogate (biomarker) endpoint



Preclinical

- In vitro experiments
- Animal models
- Research-grade assays & possibly molecular imaging

Phase 0 - I trials

- Clinical samples
- Possibly imaging – anatomic & molecular
- Early clinical assay meeting minimal analytic performance standards

Phase II - III trials

- Clinical samples
- Possibly imaging – anatomic & molecular
- Clinical assay demonstrating good analytic performance

Many factors may affect clinical assays

Example: C-reactive Protein (CRP) in cardiovascular risk assessment

Preanalytic

- Physiologic
  - Race
  - Age
  - Sex
  - Season
  - Biological variation
  - Lifestyle (exercise, smoking, obesity, alcohol, anti-inflammatory
  - drugs, hormone therapy)
  - Other (altitude, pregnancy)
- Specimen collection
  - Fasting
  - Time of collection
  - Specimen type
  - Time and temperature of storage

Analytic

- Laboratory methodology
- Detection limit
- Precision
- Antigen excess
- Matrix effects
- Calibration/curve -fitting
- Method correlation
- Reference materials
- Standardization
- Quality assessment

Clinical Chemistry 2003;49:1258–1271

Assay fit-for-purpose

Assay fit-for-purpose

- What role will the biomarker play in the drug development process?
- Does the assay measure what it is intended to measure?
- Can the assay be performed on the types of specimens available?
- Is the assay analytic performance acceptable *for the context in which the biomarker will be used*?

Description of the assay

Description of the assay & scoring system

- Analyte(s) to be measured
- Specimens needed (*in vitro*) and/or conditions of measurement (*in vivo*), including pre-analytic requirements
- Technical platform
- Sources of assay components (e.g., reagents, chips, calibrators, equipment)
- Positive & negative controls, calibrators, reference standards

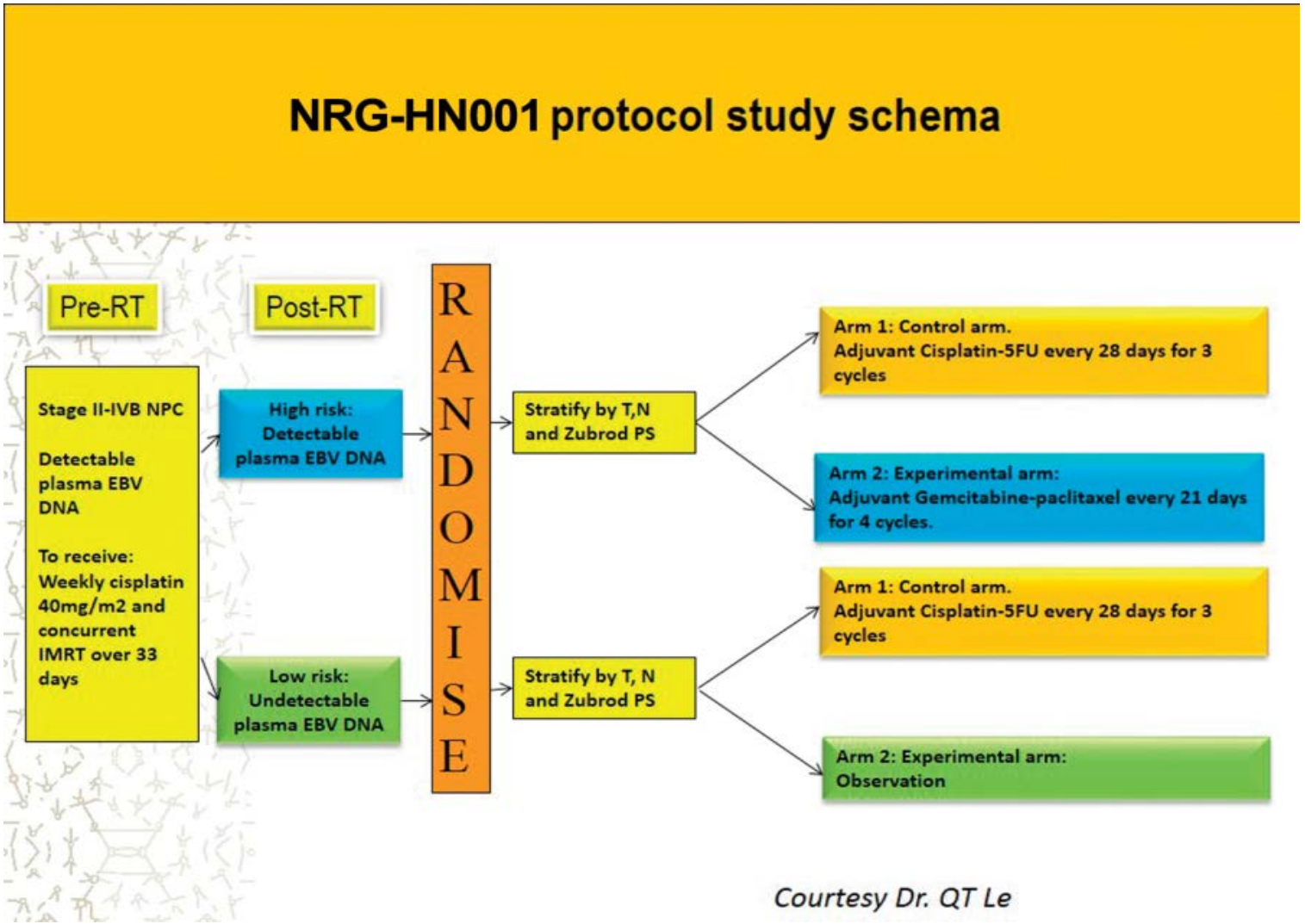
(cont.)

Description of the assay (cont.)

- Technical protocol
- Type of data
  - Quantitative/continuous (with or without cut-points)
  - Semi-quantitative
  - Qualitative/non-ordered categorical
- Scoring procedure
  - Algorithm or score calculation
  - Cut-points
- Interpretation (for clinical use)

Assay harmonization for a multi-site clinical trial

Plasma Epstein-Barr Virus (EBV) DNA as a prognostic stratifier in a treatment trial for nasopharyngeal carcinoma



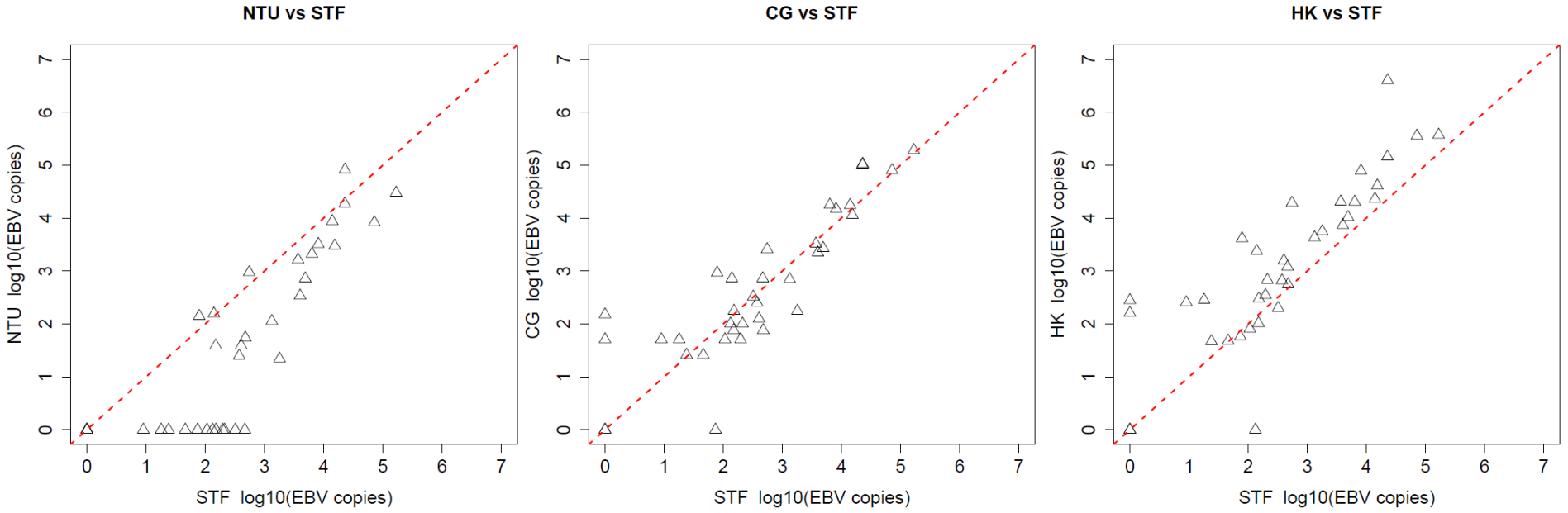
Courtesy Dr. QT Le

Impact of different assay components

An International Collaboration to Harmonize the Quantitative Plasma Epstein-Barr Virus DNA Assay for Future Biomarker-Guided Trials in Nasopharyngeal Carcinoma (4 sites: STF, NTU, CG, HK)

BEFORE HARMONIZATION

(40 samples, $\log_{10}$(EBV copies))



EBV DETECTION RATES: NTU 58%, CG 93%, HK 93%

Plots constructed from data in Clin Cancer Res 2013;19:2208-2215

Harmonization of assays

An International Collaboration to Harmonize the Quantitative Plasma Epstein-Barr Virus DNA Assay for Future Biomarker-Guided Trials in Nasopharyngeal Carcinoma

Intraclass correlation (ICC) for each site when compared to index site (STF) before and **after harmonization** of PCR master mixes and calibrators

Site	**Pre-harmonization ICC (95% CI) N = 40**	**Post harmonization ICC (95% CI) N = 10**
NTU vs. STF	0.62 (0.39-0.78)	0.83 (0.50-0.95)
CG vs. STF	0.70 (0.50-0.83)	0.95 (0.83-0.99)
HK vs. STF	0.59 (0.35-0.76)	0.96 (0.86-0.99)

Clin Cancer Res 2013;19:2208-2215

Assay performance assessment

Specimen and pre-analytic factors

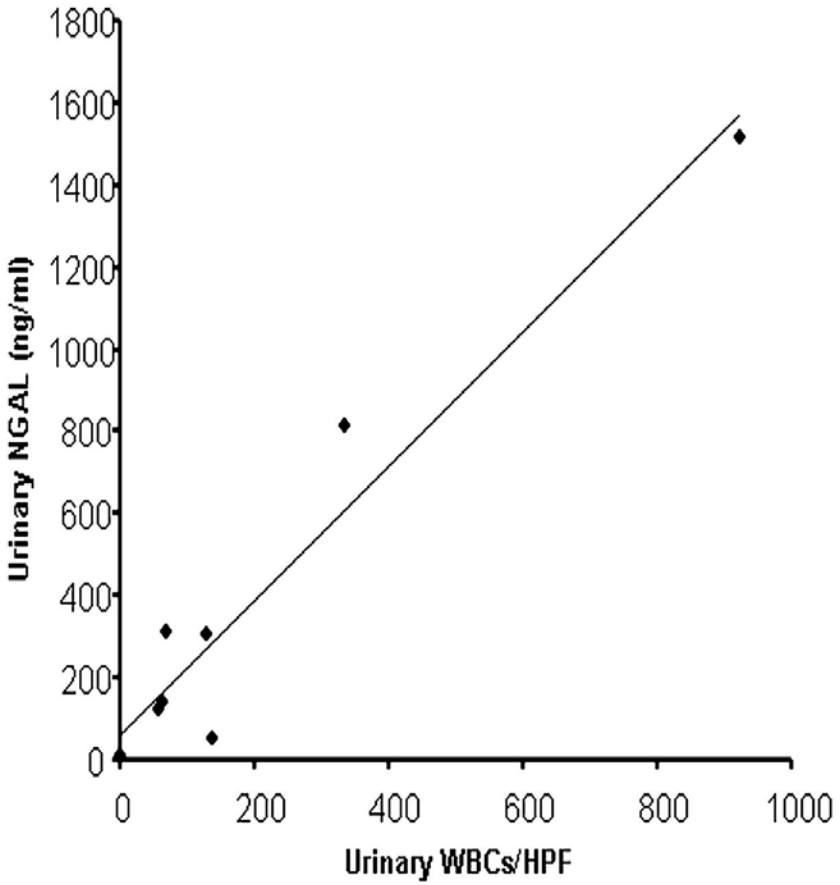
- Patient physiologic factors & state at specimen collection
- Specimen collection method, processing & storage
- Specimen quality screening
- Minimum required amount of specimen
- Feasibility of collecting needed specimens in clinical trial setting

Cancer Cytopathology 2011;119: 92-101
Arch Pathol Lab Med 2014;138:526–537

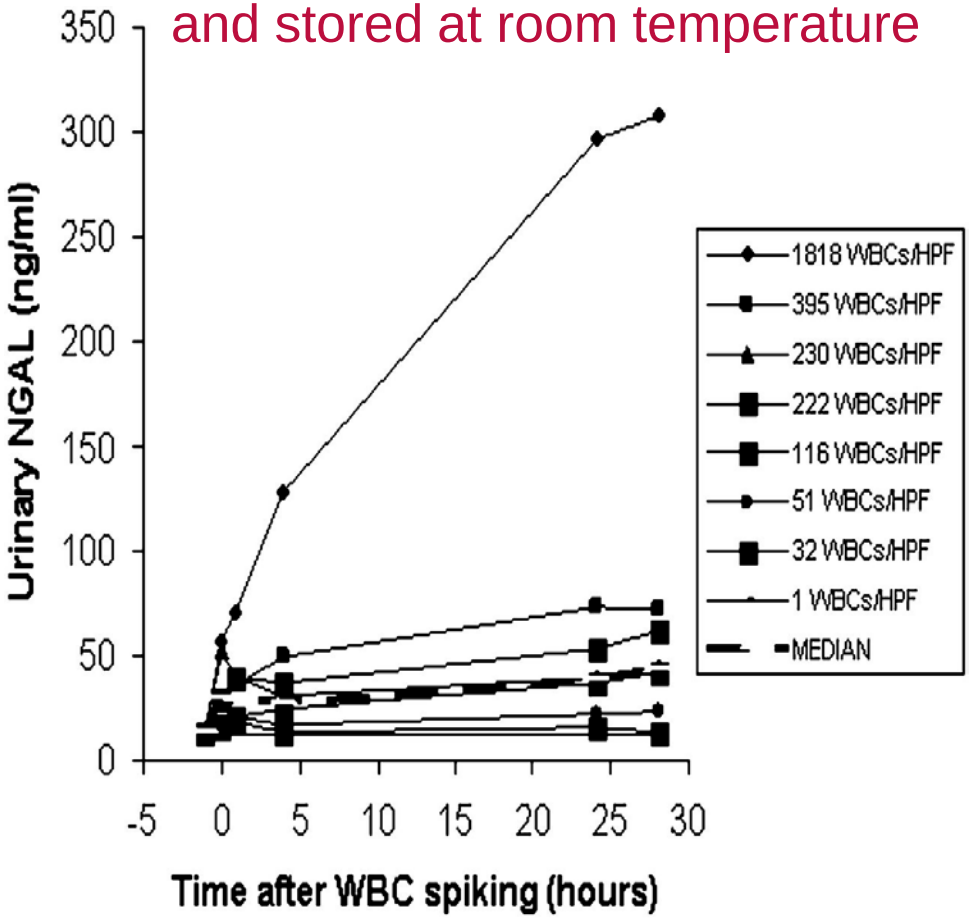
Specimen and pre-analytic factors

Urinary neutrophil gelatinase-associated lipocalin (NGAL) for early detection of acute kidney injury

Effect of WBCs on urine NGAL concentrations at baseline



Increased urinary NGAL over time in urine samples spiked with WBC and stored at room temperature



Nephrol Dial Transplant 2013;28:1175-1185

Assay analytic performance

- Precision and reproducibility
- Bias and accuracy
- Analytic sensitivity; limits of detection and quantification; linearity
- Analytic specificity
- Data to support clinical cut-off (if applicable)

Arch Pathol Lab Med 2009;133:743-755

WARNING: Literature inconsistent in use of terminology, so define terms when presenting assay performance metrics

Assay analytic validation
Precision and reproducibility

- Precision: Repeatability of measurements under essentially unchanged assay conditions in one lab (e.g., within-assay run)
- Intermediate precision: Consistency of measurements when there is variation in one or more factors (e.g., time, assay run, calibration) within a lab
- Reproducibility: Consistency of measurements between labs or under substantially different conditions such as measuring systems

Precision of a troponin T ELISA specific for cardiac troponin T isoform

Intra-assay precision and day-to-day imprecision

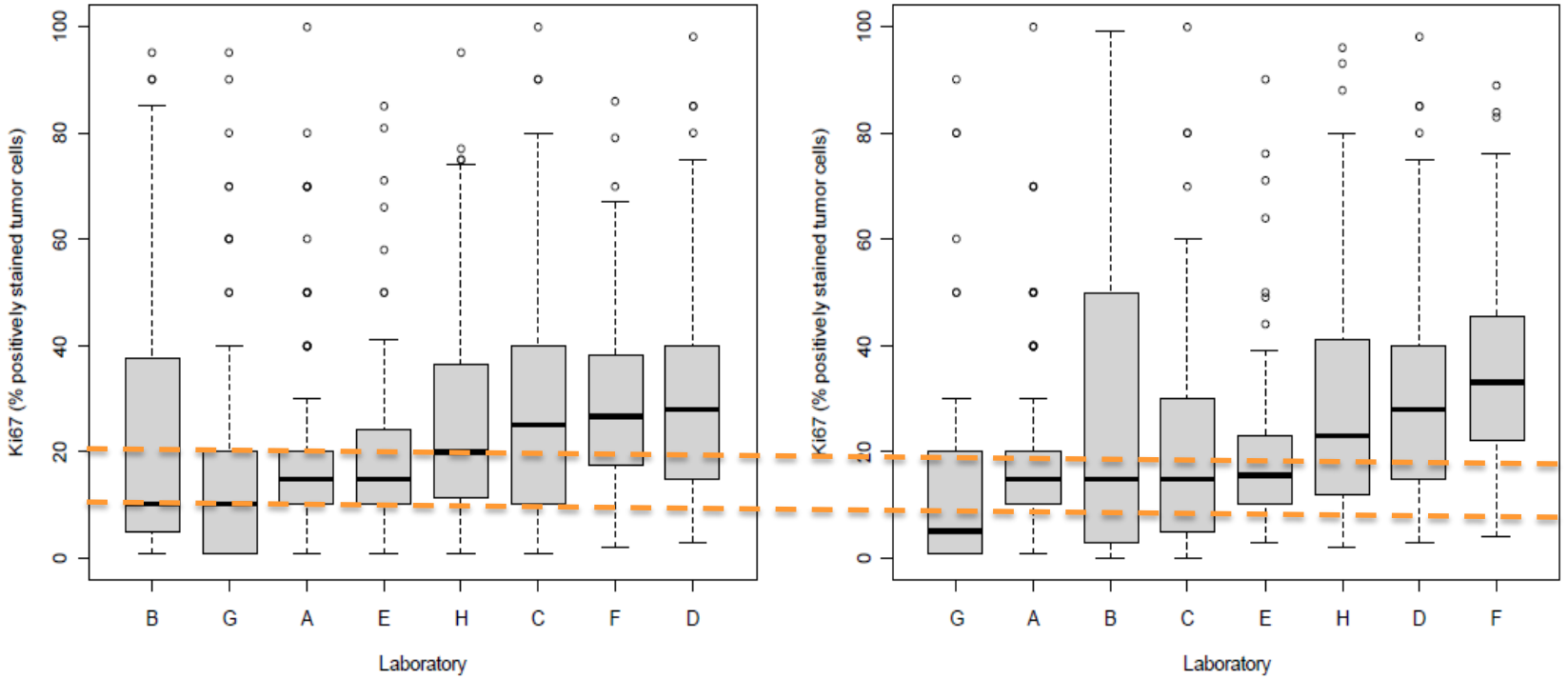
Mean ± SD concentration and CV = (SD/mean)×100% were calculated.

- Intra-assay precision measuring 10 times five serum samples with cTnT concentrations of 0.19, 5.14, 5.38, 9.39, and 13.74 μg/L
  - Mean ± SD concentrations, in μg/L (and CV), were: 0.19 ± 0.01 (4.1%), 5.14 ± 0.06 (1.3%), 5.38 ± 0.12 (2.2%), 9.39 ± 0.11 (1.2%), and 13.74 ± 0.1 (0.7%)
- Day-to-day imprecision analyzing five serum samples of different cTnT concentrations (0.19, 0.30, 0.54, 5.28, and 14.89 μg/L) once each on 10 subsequent days
  - Mean ± SD concentrations, in μg/L (and CV), were: 0.19 ± 0.01 (5.8%), 0.3 ± 0.01 (3.8%), 0.54 ± 0.02 (4.5%), 5.28 ± 0.17 (3.2%), and 14.89 ± 0.29 (2.0%)

Clinical Chemistry 1997; 43:458–466

Ki67 IHC reproducibility assessment

Boxplots of Ki67 (% positive invasive tumor cells) with 8 labs assessing different TMA sections of same set of 100 breast tumors (most clinical cut-offs in 10-20% range)



Centrally stained, locally scored
Median range: 10% to 28%
ICC: 0.71, 95% CI=(0.47,0.78)

Locally stained, locally scored
Median range: 5% to 33%
ICC: 0.59, 95% CI=(0.37,0.68)

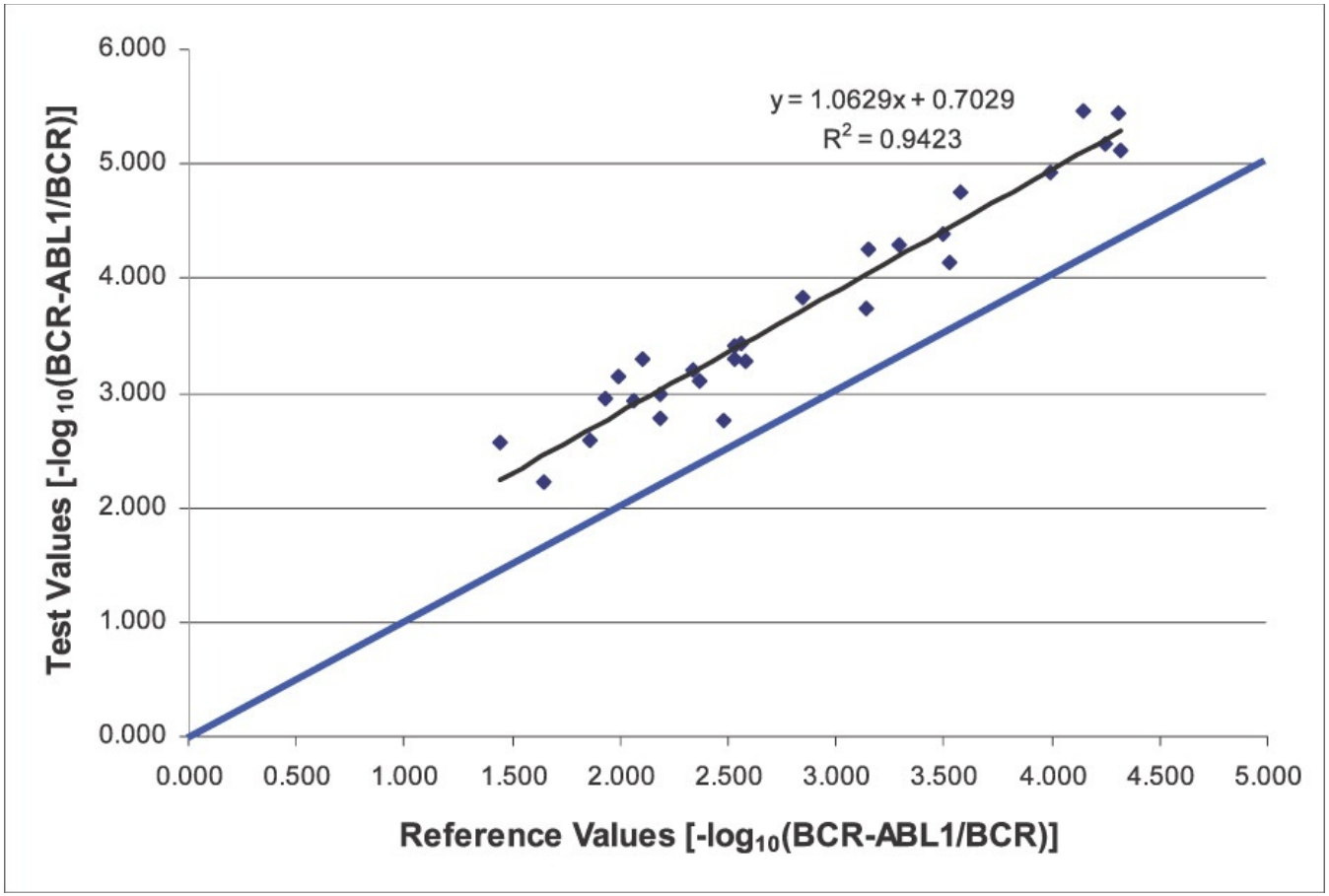
J Natl Cancer Inst 2013;105:1897-1906

Assay analytic validation
Bias & accuracy

- Bias: Amount by which an average of many repeated measurements made using the assay systematically over- or under-estimates truth (often only reference standard method result is available)
- Accuracy: Closeness of agreement between the test results obtained using the new biomarker test and results obtained using a reference standard method widely accepted as producing "truth" for the analyte
  - Reference standard method must be clearly identified (sometimes there is no real measure of "truth")
  - Incorporates elements of both bias and precision
  - Usage often ambiguous

BCR-ABL copy number assay bias assessment

*BCR-ABL*1 copy number assessed by qRT-PCR for minimal residual disease (MRD) monitoring in hematologic malignancies



- Constant bias of test relative to reference
- R^2 = 0.9423 does not capture bias

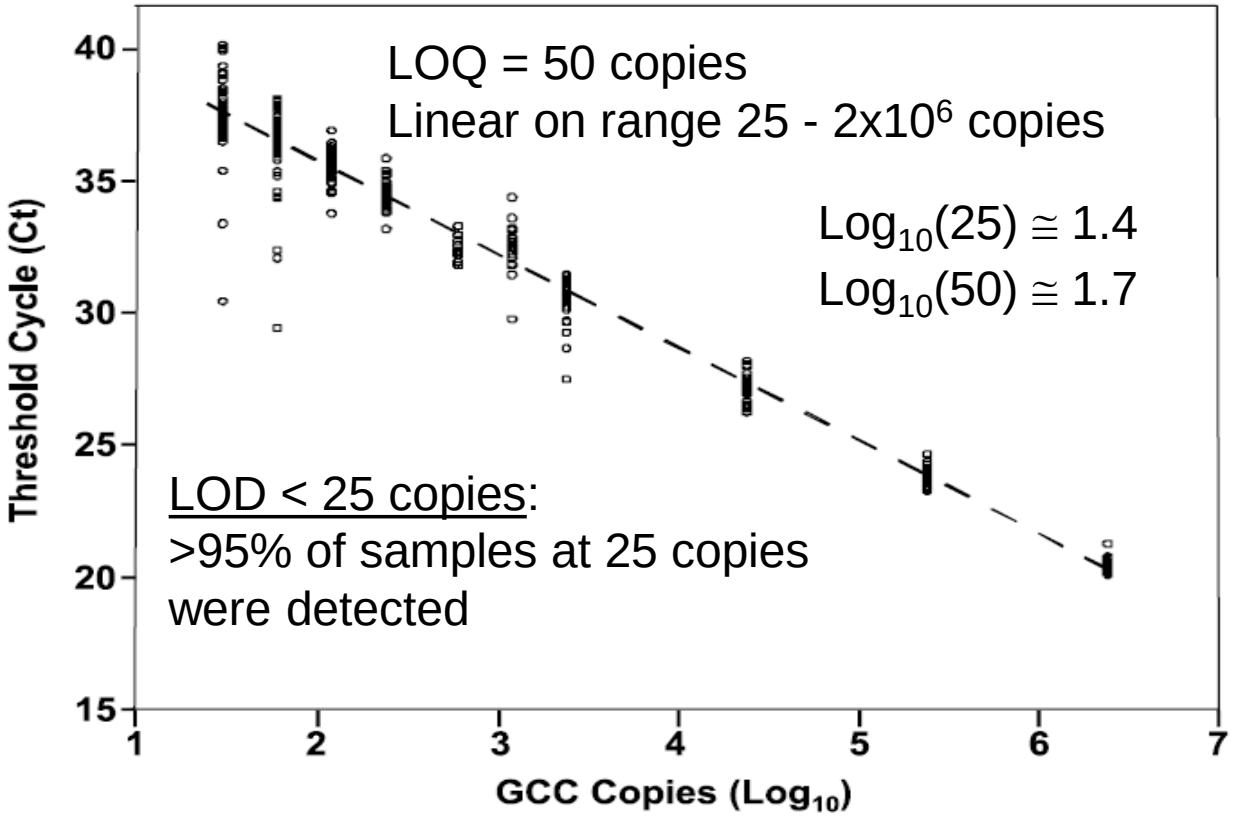
Arch Pathol Lab Med 2012;136:33–40

Assay analytic validation
Sensitivity, LOD, LOQ, and linearity

- Analytic sensitivity:
  - Binary tests: Proportion of positive tests obtained on cases that are truly positive by some reference method
  - Quantitative tests: change in the test output relative to change in the actual amount of analyte
- Limit of detection (LOD): Smallest amount of analyte detectable with specified probability
- Limit of quantitation (LOQ): Smallest amount of analyte detectable with acceptable precision & bias
- Linearity: change in test result proportional to true change in concentration

GCC mRNA qRT-PCR assay sensitivity, LOD, LOQ, and linearity assessment

Quantitative Assay to Detect Occult Micrometastases by qRT-PCR of Guanylyl Cyclase C (GCC) mRNA in Patients with Colorectal Cancer



Clin Cancer Res 2006;12:4545-4552

Assay analytic validation
Specificity

- Analytic specificity:
  - Binary tests: Proportion of negative test results obtained on cases that truly do not possess the entity or analyte of interest
  - Quantitative tests: Ability of test to accurately quantify an entity or analyte in the presence of cross-reacting or interfering substances
  - Consider that cross-reacting and interfering substances may be endogenous or exogenous

Specificity assessment for a qualitative real-time PCR high-risk test for detection of DNA from high-risk HPV types

Potential cross-reactive and interfering substances evaluated (all tested negative)

Cross-reactivity panel

41 bacteria, viruses and fungi, including 15 low-risk HPV types and other organisms that can be found in the female anogenital tract

Interfering substances panel

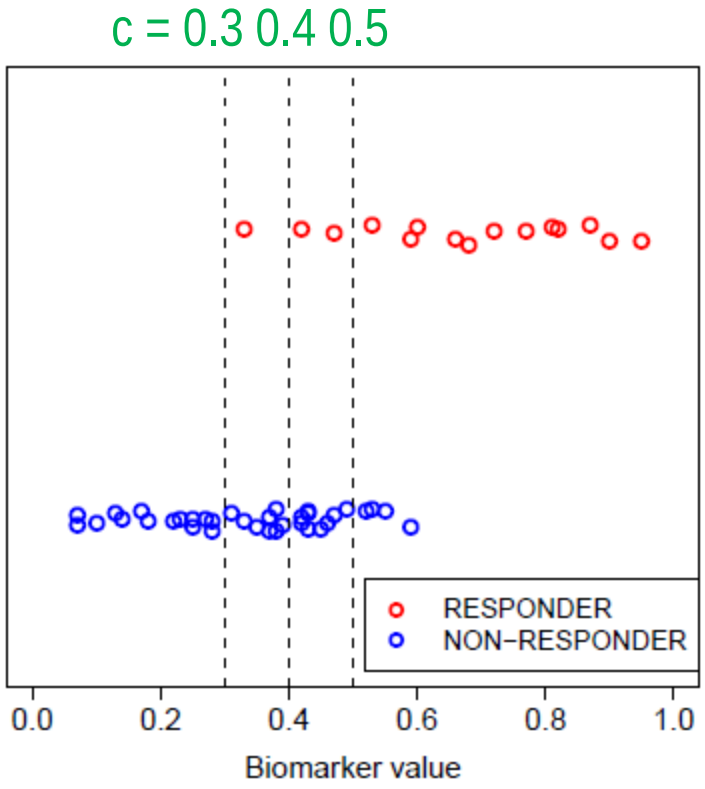
Blood
Mucous
Anti-fungal vaginal creams & suppositories
Anti-itch vaginal creams
Lubricants
Contraceptive foams
Deodorant suppositories
Douches

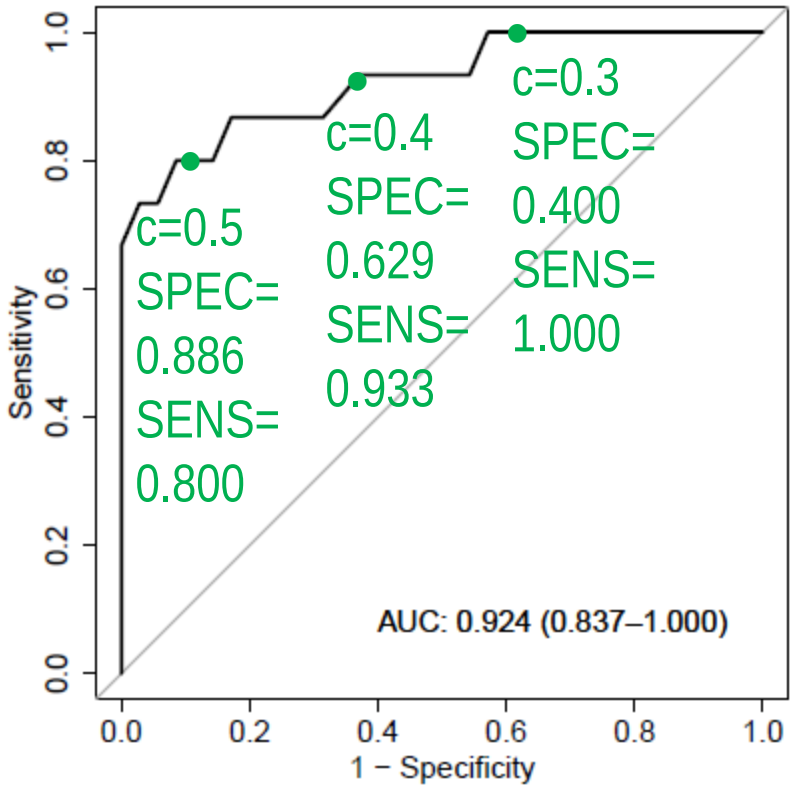
Journal of Clinical Virology 2009;45:513-517

Assay analytic validation

Data to support clinical cut-off

Receiver operating characteristic (ROC) curve plots SENSITIVITY vs. 1-SPECIFICITY to display tradeoff based on different choices of cutoff (c) applied to a continuous biomarker

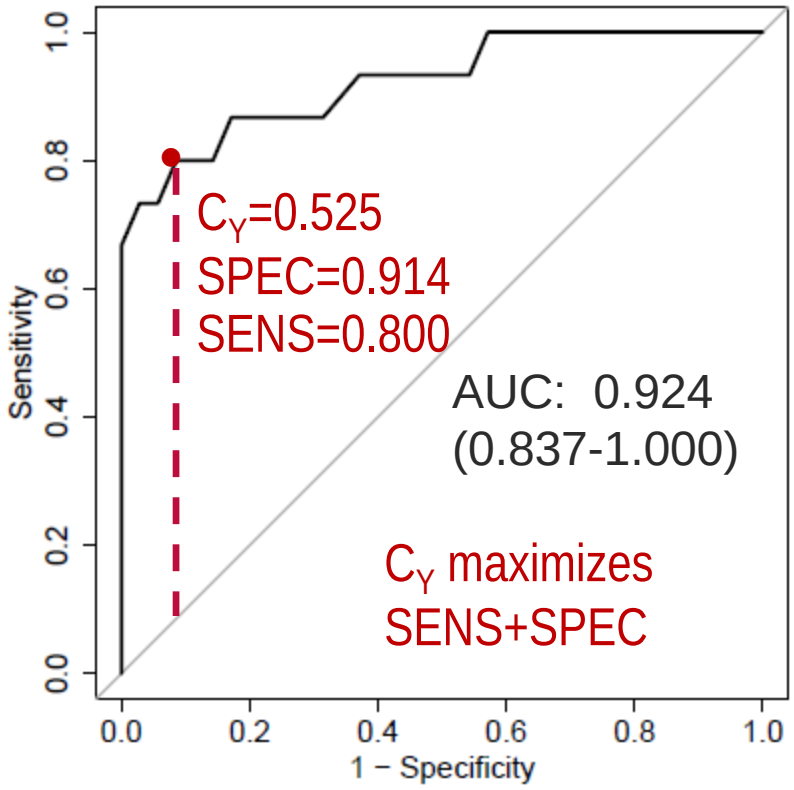




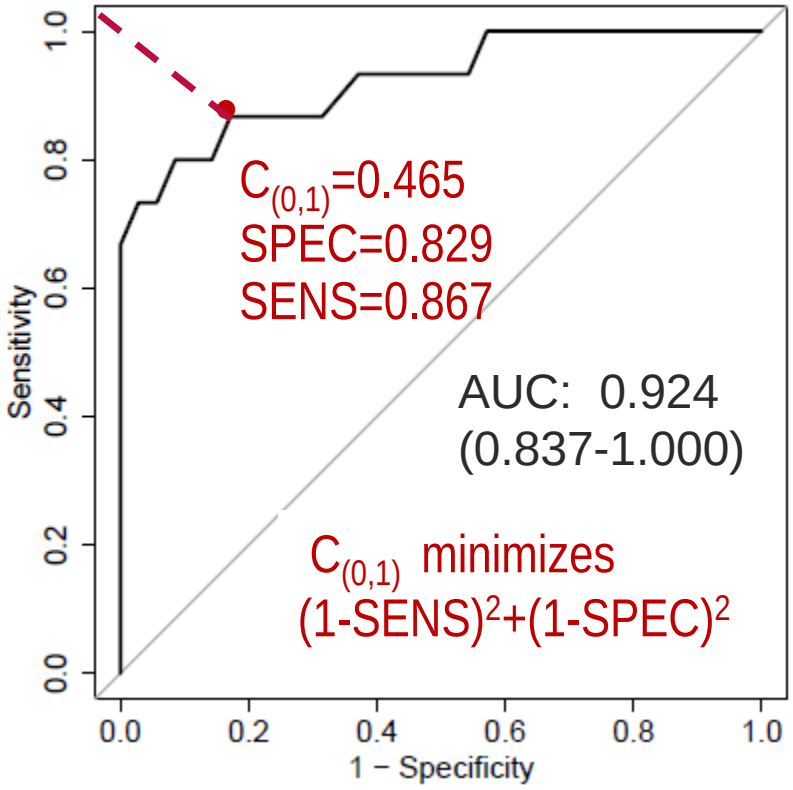
Assay analytic validation

Classical optimality criteria for cut-off selection

Youden index: Select cut-off c_Y that maximizes vertical distance from diagonal line to ROC curve



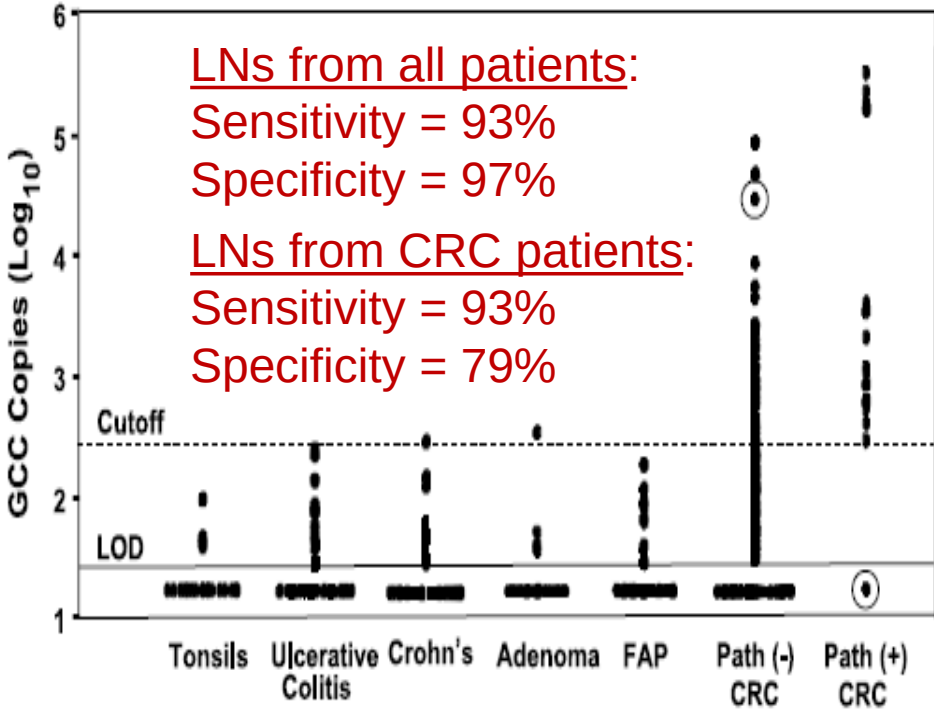
Closest to (0,1): Select cut-off $c_{(0,1)}$ that minimizes distance from point (0,1) to ROC curve

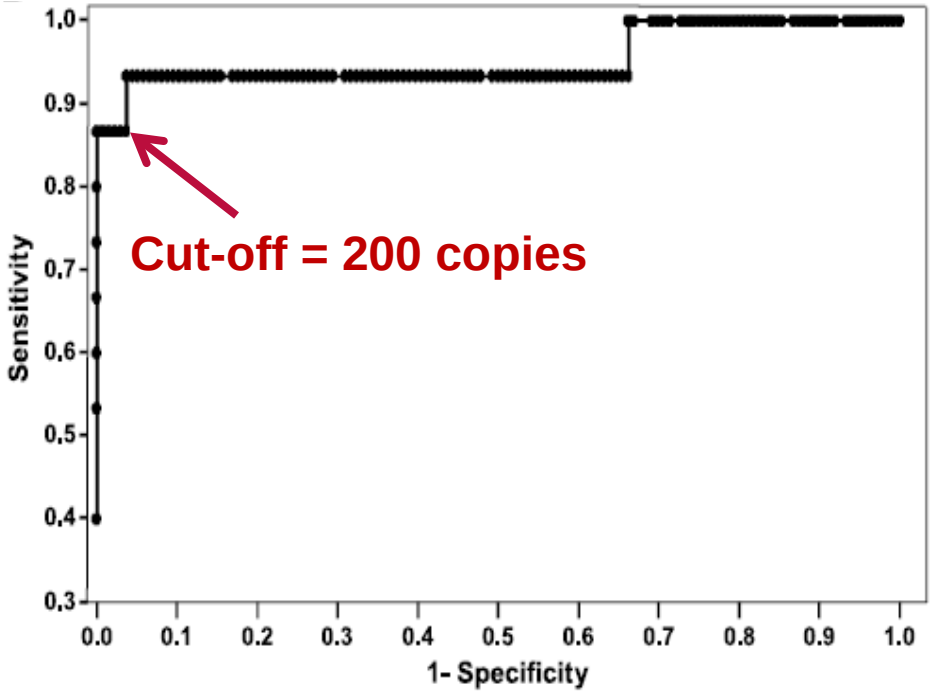


Assay analytic validation
Data to support clinical cut-off

Quantitative Assay to Detect Occult Micrometastases by qRT-PCR of Guanylyl Cyclase C (GCC) mRNA in Patients with Colorectal Cancer

Results from analyses of 546 lymph nodes (LNs) from 48 patients





Clin Cancer Res 2006;12:4545-4552

Assay analytic validation
Selection of clinical cut-off

- Additional considerations:
  - Trade-off between sensitivity and specificity will depend on risks associated with false positives and false negatives
  - “Optimized” cut-offs produce overly optimistic accuracy results and require validation with independent data
  - Cut-offs may not be transportable for assays which lack reproducibility between laboratories or scorers

Recommended references
General single analyte assays

1. Chau CH et al., Clin Cancer Res 2008;14:5967-5976
2. Jennings L et al., Arch Pathol Lab Med 2009; 133:743–755 (and other articles in that same special issue)
3. Ledue TB, Rifai N, Clinical Chemistry 2003;49:1258–1271
4. Linnet K, Boyd JC. Selection and analytical evaluation of methods with statistical techniques. In Burtis CA, Ashwood ER, Bruns DE (eds). Tietz Textbook of Clinical Chemistry and Molecular Diagnostics (5th edn). Elsevier Saunders, St Louis, MO, 2012, pp. 7–47.
5. Pennello GA, Clinical Trials 2013;10:666–676

Assorted FDA and CLSI guidance documents

Recommended references
Beyond single analytes

Next generation sequencing assays

1. Frampton G et al., Nature Biotech 2013;31:1023-1031
2. Linderman et al., BMC Medical Genomics 2014;7:20

Omics signatures/classifiers

1. IOM Report on Translational Omics (G. Omenn, Chair) (http://www.iom.edu/Reports/2012/Evolution-of-Translational-Omics.aspx)
2. McShane L et al., Nature 2013;502:317-320
3. McShane L et al., BMC Medicine 2013;11:220

THANK YOU!

NATIONAL CANCER INSTITUTE

www.cancer.gov **www.cancer.gov/espanol**