cap

Pathology and Laboratory Quality Center

Principles of Analytic Validation of Immunohistochemistry Assays

Published : *Archives of Pathology and Laboratory Medicine*

March 19, 2014

Objectives

- Apply evidence-based guidelines to ensure each Immunohistochemistry (IHC) assay is validated prior to reporting on patient samples
- Recognize the requirements for revalidation
- Understand possible differences in validation requirements based on variations in fixative or specimen type
- Understand how the quality of evidence impacts the recommendations related to the validation statements

Introduction

- Laboratories are required to validate all assays before testing patient specimens.

- There is significant variation in validation practices for IHC assays.

- Current guidelines exist only for HER2 and ER/PgR.

Background

CAP Laboratory Improvement Programs

Immunohistochemistry Validation Procedures and Practices

A College of American Pathologists Survey of 727 Laboratories

Lindsay B. Hardy, MD; Patrick L. Fitzgibbons, MD; Jeffery D. Goldsmith, MD; Richard N. Eisen, MD; Mary Beth Beasley, MD; Rhona J. Souers, MS; Raouf E. Nakhleh, MD

• *Context.*—The immunohistochemistry (IHC) laboratory represents a dynamic area of surgical pathology with limited practice guidelines. Studies have shown significant interlaboratory variability in results.

Objective.—To establish baseline parameters for IHC validation procedures and practice, and to assess their feasibility of implementation.

Design.—In September 2010, a questionnaire was distributed by the College of American Pathologists. It was composed of 32 questions relating to nonpredictive assays as well as non-US Food and Drug Administration (non-FDA)–approved, predictive IHC assays other than human epidermal growth factor 2 (HER2/*neu*).

Results.—For non-FDA approved, nonpredictive IHC assays, 68% of laboratories had a written validation procedure. Eighty-six percent of laboratories validated the most recently introduced nonpredictive antibody. Seventy-five percent used 21 or fewer total cases for the validation and 40% used weakly or focally positive cases. Forty-six percent of respondents had a written procedure for validation procedures for non-FDA approved, predictive marker IHC assays other than HER2/*neu*. Seventy-five percent of laboratories validated the most recently introduced predictive antibody other than HER2/*neu*. Fewer than half used 25 or more cases for the validation, and 47% used weakly or focally positive cases.

Conclusion.—Some laboratories have written validation procedures that appear to build upon HER2/*neu* testing guidelines. Some laboratories also manage to validate new antibodies according to those standards; however, many do not. There appears to be a need for further validation guideline development for nonpredictive and non-FDA approved predictive antibody IHC assays.

(*Arch Pathol Lab Med.* 2013;137:19–25; doi: 10.5858/arpa.2011-0676-CP)

Validation Practices - Non Predictive Factor Assays

Procedures	Yes	No
Lab has written validation procedure?	68%	28%
Procedure specifies # validation cases?	54%	44%
Procedure specifies when revalidation needed?	46%	46%
Cytology specimens addressed?	37%	63%

Hardy et al. *Arch Pathol Lab Med* 2013;137:19-25

Validation Practices - Non Predictive Factor Assays

Procedures	Yes	No
Change in antigen retrieval method?	71%	25%
Change in detection method?	74%	23%
Change in instrumentation?	74%	24%
Change in fixative?	65%	30%

Hardy et al. *Arch Pathol Lab Med* 2013;137:19-25

Introduction

- CAP convened expert and advisory panels to systematically review published data and develop evidence-based recommendations
- Closely followed IOM Clinical Practice Guidelines
  - Transparency
  - Manage conflicts of interest
  - Multidisciplinary panel
  - Patient advocate (N/A for this panel)
  - Systematic Review
  - Considered judgment

Principles of Analytic Validation for IHC Assays: Expert and Advisory Panel

Chair

Patrick Fitzgibbons, MD

Regan Fulton, MD, PhD

Jeffrey Goldsmith, MD

Thomas Haas, DO

Rouzan Karabakhtsian, MD, PhD

Patti Loykasek, HT(ASCP)QIHC

Monna Marolt, MD

Steven Shen, MD, PhD

Paul Swanson, MD

Advisory Panel Members

Raouf Nakhleh, MD, Center

Richard Brown, MD

Richard Eisen, MD

Hadi Yaziji, MD

Staff

Lisa Fatheree, SCT(ASCP)

Tony Smith, MLS

Consultant Methodologist

Linda Bradley, PhD

Systematic Evidence Review

- Identify Key Questions
- Literature search
- Data extraction
- Develop proposed recommendations
- Open comment period
- Considered judgment process

Introduction

- Overarching questions:

1. What is needed for initial analytic assay validation before placing any immunohistochemical test into clinical service ?

2. What are the revalidation requirements?

Scope Questions

1. When and how should validation assess

    - analytic sensitivity
    - analytic specificity
    - accuracy (assay concordance)
    - precision (inter-run and inter-operator variability)?

Scope Questions (cont.)

2. What is the minimum number of positive and negative cases needed to analytically validate an IHC assay for its intended use(s)?

   - Non-predictive markers
   - Predictive markers
   - Identifying infectious organisms
   - Rare antigens

Should expression levels be specified for positive cases?

Scope Questions (cont.)

3. What parameters should be specified for the tissues used in the validation set?
   - Cytology specimens
   - Minimum tissue size or minimum quantity of cells
   - Neoplastic vs non-neoplastic tissues

Scope Questions (cont.)

4. How do the following preanalytic variables influence analytic validation?
   - Type of fixative
   - Type of decalcification solution
   - Time in decalcification solution
   - Validation tissues processed in another laboratory
5. What conditions require assay revalidation?

Systematic Evidence Review

- Literature search
  - January 2004 – May 2013
  - 1,463 studies met inclusion criteria
    → Reviewed by panel
  - 126 studies identified for full data extraction

Systematic Evidence Review

- Evidence Evaluation
  - Quality (rate strength of evidence)
  - Quantity
  - Consistency

Quality Assessment

- Individual studies graded on specific criteria by the methodology consultant (LAB)
- Criteria included
  - Quality and execution of studies
  - Quantity of data (number and size of studies)
  - Consistency and generalizability of the evidence across studies
    - Adequate descriptions of the test
    - Adequate descriptions of the basis for the "right answer"
    - Reproducibility of test results
    - Avoidance of biases
    - Analysis of data

Grades for Strength of Evidence

Grade	Description
Convincing	Level 1 or 2 studies with an appropriate number and distribution of challenges and reported consistent and generalizable results.
Adequate	Level 1 or 2 studies that lacked the appropriate number and distribution of challenges OR were consistent but not generalizable.
Inadequate	Combinations of Level 1 or 2 studies that show unexplained inconsistencies, OR one or more lower quality studies (Level 3 or 4), OR expert opinion.

Level 1: Collaborative study using a large panel of well-characterized samples; summary data from external proficiency testing schemes or inter-laboratory comparisons
Level 2: High quality peer-reviewed studies
Level 3: Lower quality peer-reviewed studies OR expert panel reviewed FDA summaries
Level 4: Unpublished or non-peer reviewed data

Grades for Strength of Recommendation

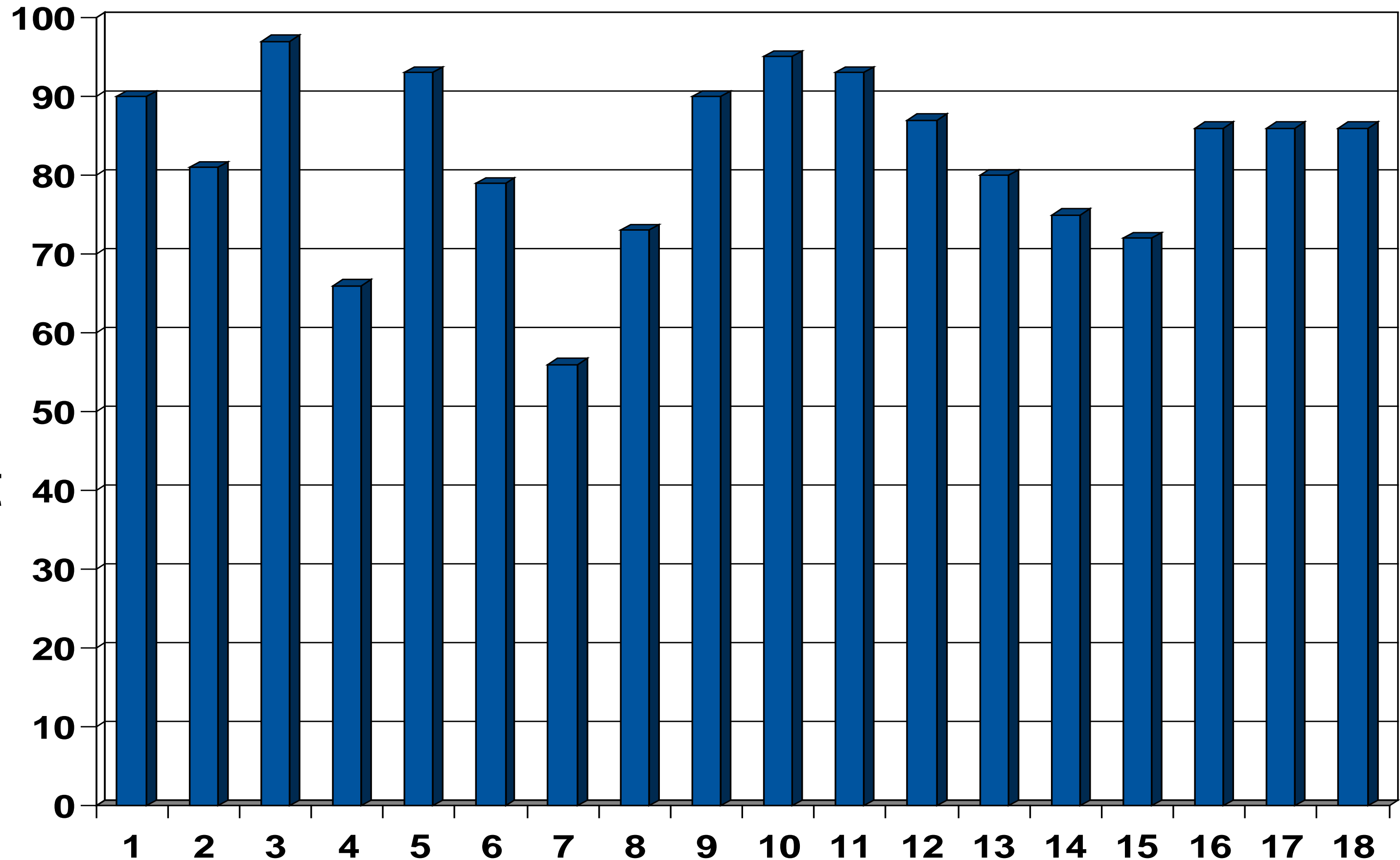
Designation	Rationale
Strong Recommendation	Strength of evidence is Convincing based on consistent, generalizable, good quality evidence; further studies are unlikely to change the conclusions
Recommendation	Strength of evidence is Adequate based on limitations in the quality of evidence; further studies may change the conclusions
Expert Consensus Opinion	Important validation element to address but strength of evidence is Inadequate; gaps in knowledge may require further studies

Systematic Evidence Review

- Open comment period (July 2013):
  - 18 draft recommendations and 5 methodology questions
  - 263 respondents; 1,037 comments

Open Comment Period

%
Agreement



Original Draft Proposed Recommendation
Final Recommendations Combined/Condensed into 14 Total

Systematic Evidence Review

- Considered judgment process
  - Panel reviews and considers
    - Feedback
    - Quality/quantity/consistency of evidence
    - Benefits/harms
    - Value versus cost/burdens
    - Regulatory requirements
    - Expert opinion
  - 14 final recommendations

ASCO/CAP HER2 Guideline Recommendations Summary of Changes

Initial Test Validation

2007	2013
25–100 samples	20(+), 20(-) for FDA-approved assays 40(+), 40(-) for LDTs Not applicable if assay was previously validated and lab has successful PT performance

ASCO/CAP HER2 Guideline Recommendations Summary of Changes

Concordance

2007	2013
If <95% for any result category, cases with that test result must be automatically reflexed to alternative method	Specific concordance requirements are not required Laboratories must comply with accreditation and PT requirements

The Guidelines

Guideline 1

Recommendation: **Laboratories must validate all immunohistochemical tests before placing into clinical service.** *Note: Such means include (but are not necessarily limited to):*

- *Correlating the new test's results with the morphology and expected results;*
- *Comparing the new test's results with the results of prior testing of the same tissues with a validated assay in the same laboratory;*
- *Comparing the new test's results with the results of testing the same tissue validation set in another laboratory using a validated assay;*
- *Comparing the new test's results with previously validated non-immunohistochemical tests; or*
- *Testing previously graded tissue challenges from a formal proficiency testing program (if available) and comparing the results with the graded responses.*

Guideline 1

- Strength of Evidence:
  - **Adequate** to support when analytic validation should be done and that it should include determination of concordance and precision
  - **Inadequate** to assess how validation should be done with regard to the listed approaches, but did show that these approaches have been used.
- Rationale: Analytic validation provides a net benefit for the overall performance and safety of IHC tests by contributing to the avoidance of potential harms related to analytic false positive and false negative test results.

Rationale 1

- Validation set should include:
  - Positive, negative, and low positive tissues
  - Should not be all normal tissues
  - Should reflect the intended use of the assay

- Positive and negative cell types on the same section could be used as separate challenges

Guideline 2

Recommendation: For initial validation of every assay used clinically (with the exception of HER2, ER and PgR, for which established validation guidelines already exist), laboratories should achieve at least 90% overall concordance between the new test and the comparator test or expected results. If concordance is less than 90%, laboratories need to investigate the cause of low concordance.

Guideline 2

- Strength of evidence
  - **Adequate** to support a 90% (versus 95%) overall concordance benchmark for analytic validation of IHC tests (except HER2, ER, PgR)
- Median overall concordance in a two-year inter-laboratory comparison of CD117 IHC and target results was 87.6% (Hsi, 2001)
- Median overall concordance in 5 comparisons of different HER2 IHC tests was 89.0% (range 74–92%), with 2 of 5 studies >90% concordant. (Boers, 2011; Mayr, 2009; Moelans, 2010; O'Grady, 2010; van der Vegt, 2009)
- Median overall concordance in 5 comparisons of HER2 IHC tests to HER2 ISH tests was 88.2% (range 66– 94%), with 2 of 5 comparisons >90% concordant (Dorfman, 2006; Jordan, 2012; Lotan, 2011; Phillips 2007)
- Median overall concordance in 6 comparisons of IHC tests (PTEN, ER, PR, HER2, MPT64, p16) to alternative referent tests (*e.g.*, RNA expression, clinical diagnosis) was 91.4% (range 74–99%), with 3 of 6 studies >90% concordant (Phillips, 2007; Baba, 2008, Lehmann-Che, 2011)

Guideline 3

Expert Consensus Opinion: For initial analytic validation of non-predictive factor assays, laboratories should test a minimum of 10 positive and 10 negative tissues. When the laboratory medical director determines that fewer than 20 validation cases are sufficient for a specific marker (e.g., rare antigen), the rationale for that decision needs to be documented.

- *Note: The validation set should include high and low expressors for positive cases when appropriate, and should span the expected range of clinical results (expression levels) for markers that are reported quantitatively.*

Guideline 3

- **Strength of Evidence**

  - **Inadequate** to support the recommended number of validation samples.

  - **Adequate** to support the distinction between non-predictive and predictive IHC tests and the use of different numbers.

Validation Using 10 and 20 Tissue Validation Sets against a 90% Concordance Benchmark

	Concordance estimate (95% CI)		
# of validation tissues	**0 discordant**	**1 discordant**	**2 discordant**
10	100% (68-100)	90% (57-100)	80% (48-95)
20	100% (81-100)	95% (75-100)	90% (69-98)

Concordance estimates with 95% confidence intervals stratified by number of observed discordant samples

Guideline 4

Expert Consensus Opinion: For initial analytic validation of all laboratory-developed predictive marker assays, laboratories should test a minimum of 20 positive and 20 negative tissues. When the laboratory medical director determines that fewer than 40 validation tissues are sufficient for a specific marker, the rationale for that decision needs to be documented.

- *Note: Positive cases in the validation set should span the expected range of clinical results (expression levels). This recommendation does not apply to any marker for which a separate validation guideline already exists.*

Guideline 4

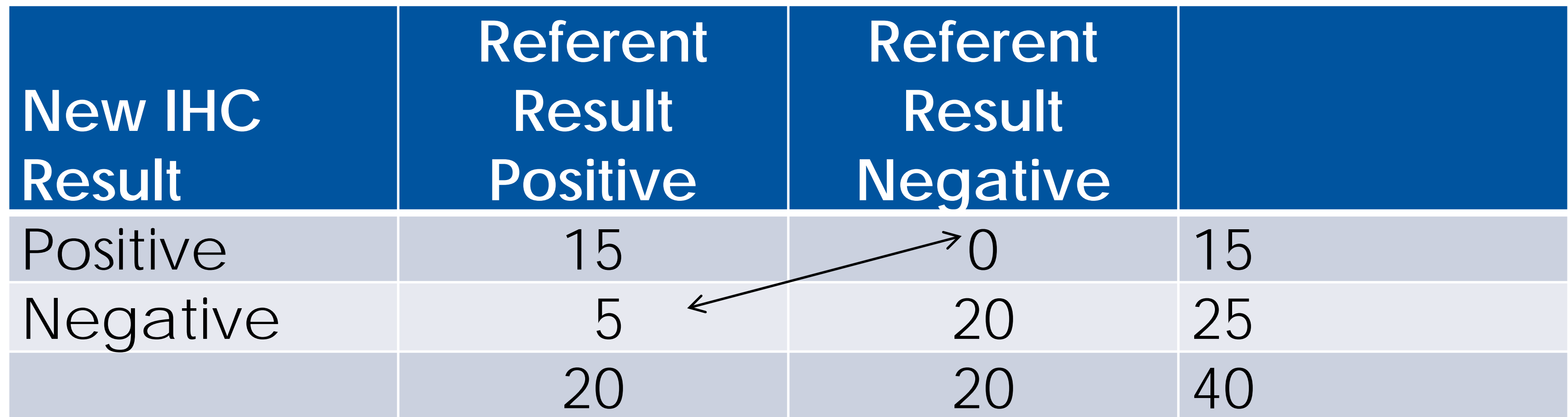
- **Strength of Evidence**
  - **Inadequate** to support the recommended number of validation samples.
  - **Adequate** to support the distinction between non-predictive and predictive IHC tests and the use of different numbers.

Validation Using a 40 Tissue Validation Set (20 Positive and 20 Negative) against a 90% Concordance Benchmark

	Concordance estimate (95% CI)				
# of validation tissues	0 discordant	1 discordant	2 discordant	3 discordant	4 discordant
20	100% (81-100)	95% (75-100)	90% (69-98)	85% (63-96)	80% (58-92)
40	100% (90-100)	97.5% (86-100)	95% (83-99)	92.5% (79-98)	90% (76-97)

Concordance estimates with 95% confidence intervals stratified by number of observed discordant samples

2x2 contingency table of a 40 tissue validation set that did not meet the benchmark (results entered into a 2x2 contingency table) with associated statistical tests



New IHC Result	Referent Result Positive	Referent Result Negative	
Positive	15	0	15
Negative	5	20	25
	20	20	40

Overall concordance: 35/40=87.5% (does not meet 90% benchmark)
Kappa: 0.75
McNemar's p: 0.13, not significant
Positive concordance: 15/20 = 75%
Negative concordance: 20/20 = 100%

Guideline 5

***Recommendation:* For a marker with both predictive and non-predictive applications, laboratories should validate it as a predictive marker if it is used as such**

- Strength of evidence:
  - Adequate to support the use of the higher validation standard (*e.g.*, number of samples) in the case of a marker with both non-predictive and predictive intended uses.

Guideline 6

Recommendation: **When possible, laboratories should use validation tissues that have been processed using the same fixative and processing methods as cases that will be tested clinically.**

- Strength of evidence
  - **Adequate** to support that laboratories should, whenever possible, use the same fixative and processing methods as cases tested clinically, in order to validate using representative specimens.

Guideline 6

- Can be difficult in reference laboratories that receive tissues with disparate fixation protocols
- Focused validation with a small number of markers may be appropriate

Guideline 7

Expert Consensus Opinion: If IHC is regularly done on cytologic specimens that are not processed in the same manner as the tissues used for assay validation (e.g., alcohol-fixed cell blocks, air-dried smears, formalin post-fixed specimens), laboratories should test a sufficient number of such cases to ensure that assays consistently achieve expected results. The laboratory medical director is responsible for determining the number of positive and negative cases and the number of predictive and non-predictive markers to test.

Guideline 7

- Strength of evidence
  - Inadequate to address the criteria and number of samples needed for validation with cytology specimens.
- Focused validation on representative antibodies used on cytologic specimens would be appropriate
- A disclaimer in the report (especially in the case of negative results) may be appropriate if assays cannot be feasibly validated:
  - "Immunohistochemistry on cytologic specimens has not been sufficiently validated; these results should be interpreted with caution."

Guideline 8

Expert Consensus Opinion: If IHC is regularly done on decalcified tissues, laboratories should test a sufficient number of such tissues to ensure that assays consistently achieve expected results. The laboratory medical director is responsible for determining the number of positive and negative tissues and the number of predictive and non-predictive markers to test.

Guideline 8

- Strength of evidence:
    - **Inadequate** to address the criteria and number of samples needed for validation with decalcified specimens.
- Focused validation on representative antibodies used on decalcified specimens would be appropriate
- A disclaimer in the report (especially in the case of negative results) may be appropriate if assays cannot be feasibly validated (ANP.22985)

Guideline 9

Recommendation: Laboratories may use whole sections, tissue microarrays (TMAs) and/or multitissue blocks (MTBs) in their validation sets as appropriate. Whole sections should be used if TMAs/MTBs are not appropriate for the targeted antigen or if the laboratory medical director cannot confirm that the fixation and processing of TMAs/ MTBs is similar to clinical specimens.

Guideline 9

- Strength of evidence
  - **Adequate** to support TMA usage; however there are many variables to be considered and thorough validation is needed for each marker.
  - **Inadequate** to recommend the *routine* use of TMA samples.
- TMAs / MTBs can be very useful in many circumstances. Beware of:
  - Proteins with high levels of heterogeneity (gastric Her2)
  - Limited tissue expression (e.g. bcl-6)

Revalidation Secondary to Assay Modification

Antibody Specific

1. Least:
   - New antibody Lot
2. Moderate:
   - Antibody dilution
   - Antibody vendor (same clone)
   - Antibody incubation or antigen retrieval times (same A.R. method)
3. Most:
   - New antibody clone

All Assays (one tier):

- Fixative type
- Antigen retrieval method
  - pH change
  - buffer type
  - heat type
- Antigen detection system
- Tissue processing equipment
- Environmental conditions
  - location
  - water supply

Evidence for Revalidation Guidelines 10-13

- Strength of evidence
  - **Inadequate** to address conditions requiring assay revalidation and whether revalidation should be the same as initial validation.

Guideline 10

Expert Consensus Opinion: **When a new reagent lot is placed into clinical service for an existing validated assay, laboratories should confirm the assay's performance with at least 1 known positive case and 1 known negative case.**

- Laboratories may want to include low-expressors, especially with predictive markers

Guideline 11

Expert Consensus Opinion: Laboratories should confirm assay performance with at least 2 known positive and 2 known negative cases when an existing validated assay has changed in any one of the following ways:

- Antibody dilution
- Antibody vendor (same clone)
- Incubation or retrieval times (same method)

- Laboratories may want to include low-expressors, especially with predictive markers

Guideline 12

Expert Consensus Opinion: Laboratories should confirm assay performance by testing a sufficient number of cases to ensure that assays consistently achieve expected results when any of the following have changed:

- Fixative type
- Antigen retrieval method (e.g., change in pH, different buffer, different heat platform)
- Antigen detection system
- Tissue processing or testing equipment
- Environmental conditions of testing (e.g. laboratory relocation)
- Laboratory water supply

Guideline 12

- **The laboratory medical director is responsible for determining how many predictive and non-predictive markers and how many positive and negative tissues to test.**

  - Reasonable approach:
    - Selection of antibodies from menu with:
      - Variable clinical uses (predictive and non-predictive)
      - Variable antigen localizations
      - Variable antibody types (monoclonal / polyclonal, etc.)

Guideline 13

Expert Consensus Opinion: Laboratories should run a full revalidation (equivalent to initial analytic validation) when the antibody clone is changed for an existing validated assay.

Guideline 14

Expert Consensus Opinion: The laboratory must document all validations and verifications in compliance with regulatory and accreditation requirements.

Summary

- Physicians and patients rely on accurate diagnostic and prognostic testing in the clinical laboratory.
- Analytic validation is essential to ensuring that an assay performs as expected, accurately identifies and/or quantifies the targeted analyte, and minimizes the chances of false positive or false negative results.
- Established guidelines are important to improve the reproducibility and consistency of the test results.

References

Early Online Release March 19, 2014

Archives of Pathology and Laboratory Medicine

http://www.archivesofpathology.org/doi/pdf/10.5858/arpa.2013-0610-CP

Disclaimer
IHC Validation Teaching PowerPoint Copyright

Effective March 19, 2014

Copyright of the line-by-line text and the teaching PowerPoint of the Principles of Analytic Validation of Immunohistochemistry Assays belongs to CAP.

Permission to reprint manuscript guidelines text for any purpose (e.g., educational or commercial) requires written permission by *Archives*

The guideline recommendations must be reproduced without modification, edits or changes to text.