

Quantitative Image Analysis of Human Epidermal Growth Factor Receptor 2 Immunohistochemistry for Breast Cancer

Guideline From the College of American Pathologists

Marilyn M. Bui, MD, PhD; Michael W. Riben, MD; Kimberly H. Allison, MD; Elizabeth Chlipala, BS, HTL(ASCP)QIHC; Carol Colasacco, MLIS, SCT(ASCP); Andrea G. Kahn, MD; Christina Lacchetti, MHS; Anant Madabhushi, PhD; Liron Pantanowitz, MD; Mohamed E. Salama, MD; Rachel L. Stewart, DO, PhD; Nicole E. Thomas, MPH, CT(ASCP)^{cm}; John E. Tomaszewski, MD; M. Elizabeth Hammond, MD

• **Context.**—Advancements in genomic, computing, and imaging technology have spurred new opportunities to use quantitative image analysis (QIA) for diagnostic testing.

Objective.—To develop evidence-based recommendations to improve accuracy, precision, and reproducibility in the interpretation of human epidermal growth factor receptor 2 (HER2) immunohistochemistry (IHC) for breast cancer where QIA is used.

Design.—The College of American Pathologists (CAP) convened a panel of pathologists, histotechnologists, and computer scientists with expertise in image analysis, immunohistochemistry, quality management, and breast pathology to develop recommendations for QIA of HER2 IHC in breast cancer. A systematic review of the literature

was conducted to address 5 key questions. Final recommendations were derived from strength of evidence, open comment feedback, expert panel consensus, and advisory panel review.

Results.—Eleven recommendations were drafted: 7 based on CAP laboratory accreditation requirements and 4 based on expert consensus opinions. A 3-week open comment period received 180 comments from more than 150 participants.

Conclusions.—To improve accurate, precise, and reproducible interpretation of HER2 IHC results for breast cancer, QIA and procedures must be validated before implementation, followed by regular maintenance and ongoing evaluation of quality control and quality assurance. HER2 QIA performance, interpretation, and reporting should be supervised by pathologists with expertise in QIA.

(*Arch Pathol Lab Med.* doi: 10.5858/arpa.2018-0378-CP)

Accepted for publication November 16, 2018.

Supplemental digital content is available for this article. See text for hyperlink.

From the Department of Anatomic Pathology, H. Lee Moffitt Cancer Center, Tampa, Florida (Dr Bui); the Department of Pathology, University of Texas MD Anderson Cancer Center, Houston (Dr Riben); the Department of Pathology, Stanford University Medical Center, Stanford, California (Dr Allison); Premier Laboratory, Longmont, Colorado (Ms Chlipala); Surveys (Mses Colasacco and Thomas), College of American Pathologists, Northfield, Illinois; the Department of Pathology, University of South Alabama, Mobile (Dr Kahn); Policy and Advocacy, American Society of Clinical Oncology, Alexandria, Virginia (Ms Lacchetti); the Department of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio (Dr Madabhushi); the Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania (Dr Pantanowitz); the Department of Pathology, University of Utah/ARUP Laboratories Inc, Salt Lake City (Dr Salama); the Department of Pathology, University of Kentucky, Lexington (Dr Stewart); the Department of Pathology and Anatomical Sciences, University at Buffalo, State University of New York, Buffalo (Dr Tomaszewski); and the Department of Pathology, University of Utah School of Medicine and Intermountain Healthcare, Salt Lake City (Dr Hammond).

Authors' disclosures of potential conflicts of interest and author contributions are found in the Appendix at the end of this article.

Corresponding author: Marilyn M. Bui, MD, PhD, H. Lee Moffitt Cancer Center, Department of Anatomic Pathology, MCC Lab, 12902 USF Magnolia Dr, Tampa, FL 33612-9416 (email: marilyn.bui@moffitt.org).

In the United States, breast cancer is the most common type of cancer in women and 1 in 8 women will develop breast cancer in her lifetime.¹ For breast cancer, the testing of prognostic and predictive biomarkers, including human epidermal growth factor receptor 2 (HER2), is the standard of care. HER2 status is determined in all patients with invasive breast cancer and provides important information to guide appropriate clinical management. Recognizing the importance of HER2 testing, the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) developed recommendations to address issues related to testing, interpretation, and reporting of HER2.² According to the ASCO/CAP guideline, quantitative image analysis (QIA) can be used to achieve consistent interpretation; however, no further information was provided to explain how QIA should be conducted, prompting the development of this guideline specific for QIA.

Advancements in genomic, computing, and imaging technology have spurred new opportunities to use QIA in diagnostic testing. A 2016 CAP survey of 826 laboratories enrolled in the Histology Quality Improvement (CAP

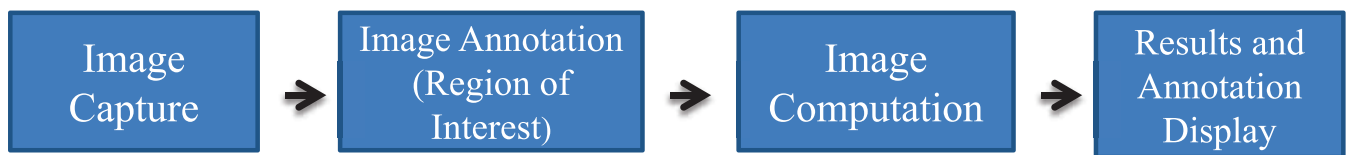


Figure 1. Basic steps involved in the quantitative image analysis test process.

HQIP-A mailing) showed that 183 (22.15%) reported using QIA.³ Notwithstanding, the implementation of QIA introduces many challenges: the initial cost of the system, changes in workflow, additional administrative burden, and training of laboratory personnel. Additionally, there are no practice guidelines to help laboratories ensure accuracy and consistency of their QIA results. To address the lack of practice guidelines, the CAP appointed an expert and an advisory panel to formulate key questions and develop recommendations. The target audience for this guideline includes laboratories that currently use or are considering the use of QIA for HER2 immunohistochemistry (IHC) for diagnostic purposes.

Quantitative image analysis is a process whereby quantitative and meaningful information is acquired from the digital images of a specimen derived from slides. When a specimen derived from slides has been scanned and digitalized, a computer algorithm identifies and helps analyze images by producing quantifiable information that enables a user to arrive at an assessment. The complete QIA process includes preanalysis steps (such as image specimen processing), imaging acquisition (scanning or slide digitalizing, image analysis, result generation), and postanalysis steps (verification, evaluation, and reporting). Refer to Figure 1 for the basic steps involved in the QIA test process.

The quality of the QIA data can be affected by many different variables including (1) preanalytic variables such as tissue handling (collection, fixation, processing), slide preparation (labeling, tissue positioning, tissue thickness, artifacts), stain variation (eg, color, platform), calibration, and image acquisition (different scanners, file format, image magnification, compression); (2) analytic variables such as algorithm choice, region of interest (ROI) selection, tumor heterogeneity (hot spots), and artifacts (tissue folds, crushed or overlapping cells); and (3) postanalytic variables such as reconciliation of discrepancies, reporting preference, and storage.^{4,5} This guideline mainly focuses on the analytic and postanalytic components of the image analysis practice.

METHODS

This evidence-based guideline has been developed by following the standards developed by the National Academy of Medicine, formerly the Institute of Medicine.⁶ A detailed description of the methods and the systematic review used to create this guideline can be found in the [supplemental digital content](#).

Panel Composition

The CAP convened an expert panel consisting of 9 practicing pathologists, 1 histotechnologist, 1 computer scientist researcher with expertise in computational imaging, and a research methodologist consultant to develop this guideline. Each expert panel member has an average of 11 years using QIA for both diagnostic and research purposes. The expert panel members also have expertise in IHC, quality management, and breast pathology. An advisory panel consisting of 4 pathologists assisted the expert panel at specific times in the development of the guideline. All panel

members, except for the methodologist consultant, volunteered their time and were not compensated for their involvement.

Conflict of Interest Policy

In accordance with the CAP conflict of interest policy (in effect April 2010), members of the expert panel disclosed all financial interests from 12 months before appointment through the development of the guideline. Individuals were instructed to disclose any relationship that could be interpreted as constituting an actual, potential, or apparent conflict. Complete disclosures of the expert panel members are listed in Appendix 1. The CAP provided funding for the administration of the project; no industry funding was involved in any aspect of the development of this guideline. See the supplemental digital content for complete information about the conflict of interest policy.

Objective

The scope of the expert panel was to provide recommendations for improving precision and accuracy in the interpretation of HER2 IHC where QIA is used. The target audience for this guideline is anatomic pathologists performing or considering QIA for diagnostic purposes, laboratory directors, and laboratory technicians/technologists in anatomic pathology. Secondary audiences include breast pathologists. The expert panel formulated and considered the following key questions:

1. What QIA system validation and daily performance monitoring is needed?
2. How does one select or develop an appropriate QIA algorithm for interpretation?
3. How does one determine the performance of a QIA algorithm?
4. What training of staff and pathologists is required and what are the competency assessments needed over time?
5. How should the results of HER2 QIA be reported?

For the purposes of this guideline, QIA is only discussed in the context of HER2 testing.

Literature Search and Collection

The expert panel developed recommendations based on evidence identified through a systematic literature review composed of database searches using Ovid MEDLINE (Ovid Technologies Inc, New York City, New York), PubMed (National Library of Medicine, Bethesda, Maryland), and Scopus (Elsevier BV, Amsterdam, The Netherlands) for articles published from January 1, 2006, through March 21, 2016. Database searches were supplemented by additional searches for indexed and unindexed (grey) literature, expert panel members were polled for known unpublished or published studies of interest not already identified, and the reference lists of included studies were scanned for relevant reports. Additional detail about the literature search and review including database search strings is available in the digital supplement content (see Supplemental Figures 1 and 2).

All identified articles were added to an online tool to manage and conduct systematic reviews, namely, DistillerSR (Evidence Partners, Ottawa, Canada). At least 2 expert panel members reviewed each article. Selection at all levels was based on predetermined inclusion/exclusion criteria.

Studies were included if they addressed QIA, were focused on surgical pathology samples from the breast, addressed HER2

Table 1. Summary of the Guideline Statements

Guideline Statement	Strength of Recommendation
1. Laboratories that choose to implement QIA for HER2 immunohistochemistry interpretation for clinical testing should select a QIA system that is validated for diagnostic interpretation. The final reporting schema should be consistent with the American Society of Clinical Oncology and the College of American Pathologists guideline “Recommendations for Human Epidermal Growth Factor 2 Testing in Breast Cancer.”	Expert consensus opinion
2. Laboratories should validate their QIA results for clinical use by comparing them to an alternative, validated method(s) such as HER2 fluorescence in situ hybridization or consensus images for HER2 immunohistochemistry.	Recommendation
3. Laboratories should ensure that the results produced by a QIA system are reproducible within and between different batch analyses.	Recommendation
4. Laboratories should ensure that the results produced by a QIA system are reproducible between operators when they select ROIs for analysis and/or perform annotation.	Recommendation
5. Laboratories should monitor and document the performance of their QIA system.	Recommendation
6. Laboratories should have procedures in place to address changes to the QIA system that could impact clinical results.	Recommendation
7. The pathologist should document that results were obtained by using QIA in the pathology report.	Expert consensus opinion
8. Personnel involved in the QIA process should be trained specifically in the use of the technology.	Recommendation
9. Laboratories should retain QIA results and the algorithm metadata in accordance with local requirements and applicable regulations.	Expert consensus opinion
10. The pathologist who oversees the entire HER2 QIA process used for clinical practice should have appropriate expertise in this area.	Recommendation
11. The pathologist finalizing the case should be knowledgeable in the use of the HER2 QIA system and visually verify that the correct ROI was analyzed, the algorithm annotated image produced, and the image analysis results.	Expert consensus opinion

Abbreviations: HER2, human epidermal growth factor receptor 2; QIA, quantitative image analysis; ROI, region of interest.

testing using IHC, and were comparative studies or guidelines, protocols, or standards. Studies were excluded if they were animal studies, meeting abstracts, noncomparative studies, or published in a non-English language.

Quality Assessment

An assessment of study quality was performed for all fully published studies meeting inclusion criteria by the research methodologist. Studies available only in abstract form did not undergo formal quality assessment. Formal quality assessment involved determining the risk of bias by assessing key indicators based on study design and methodologic rigor. Refer to the supplemental digital content for further details.

Assessing the Strength of Recommendations

Development of recommendations required that the expert panel review the identified evidence and make a series of key judgments (see Supplemental Tables 1 through 3). Grades for strength of recommendations were developed by the CAP Pathology and Laboratory Quality Center and are described in the supplemental digital content.

RESULTS

Of the 391 unique studies identified in the systematic review, 70 met the inclusion criteria and underwent data extraction. Thirty-nine of these studies made up the evidentiary base. Upon further discussion by the expert panel, only 8 directly informed the guideline statements (recommendations). All 8 were published, peer-reviewed publications and underwent data extraction and qualitative assessment (see Supplemental Figure 2).

The expert panel met 21 times, using Web-based meeting forums from August 4, 2015, through November 29, 2017. Additional work was completed via electronic mail and in 2 in-person meetings. A public comment period was held from March 6 through March 27, 2017, during which the recommendations were posted on the CAP Web site. The

expert panel agreed on the final recommendations via a formal vote.

An independent review panel masked to the expert panel and vetted through the conflict of interest process provided final approval of the guideline on behalf of the CAP Council on Scientific Affairs. The final recommendations are summarized in Table 1 and in addition to the rationale for the guideline statements below, the discussion of the benefits and harms of the guideline statements are included in the supplemental digital content.

Guideline Statements

1. Expert Consensus Opinion.—Laboratories that choose to implement QIA for HER2 IHC interpretation for clinical testing should select a QIA system that is validated for diagnostic interpretation. The final reporting schema should be consistent with the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) guideline “Recommendations for Human Epidermal Growth Factor 2 Testing in Breast Cancer.”

There were insufficient published data to inform this recommendation; however, regulatory requirements, extensive clinical experience, and a strong expert consensus were deemed *adequate* to support that a QIA system be validated for diagnostic interpretation, but *inadequate* to support that the final reporting schema be consistent with the ASCO/CAP guideline.

The goal of QIA is to detect and quantify HER2 membranous immunohistochemical staining of invasive breast cancer cells, and to provide an accurate, precise and reproducible quantitative HER2 result. To achieve this, QIA system selection and validation are paramount. Not all algorithms are designed to specifically quantify the correct staining or even provide results with the required data output. Therefore, for clinical use, laboratories should favor using a US Food and Drug Administration (FDA)–approved

Table 2. Validation of Quantitative Image Analysis (QIA) Systems

Validation Element	FDA HER2 QIA Approved (Premarket Approved) System	FDA HER2 QIA Cleared (510(k)) System	HER2 QIA Laboratory-Developed Test
QIA HER2 IHC validation	20 known positive and 20 negative HER2 IHC case image sets with known <i>HER2</i> ISH amplification status	20 known positive and 20 negative HER2 IHC case image sets with known <i>HER2</i> ISH amplification status if test is unmodified from FDA-approved version; if modified, 40 known HER2-positive and 40 known HER2-negative samples are required	40 known positive and 40 negative HER2 IHC case image sets with known <i>HER2</i> ISH amplification status
Standard	HER2 IHC method used in original algorithm	HER2 IHC method used in original algorithm; if another one is used, it must be validated against the original	HER2 IHC method used in original algorithm; if another one is used, it must be validated against the original
Validation agreement threshold	90% for HER2-positive samples, 95% for HER2-negative samples	90% for HER2-positive samples, 95% for HER2-negative samples	90% for HER2-positive samples, 95% for HER2-negative samples

Abbreviations: FDA, US Food and Drug Administration; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; ISH, in situ hybridization.

system and/or algorithms and avoid those not intended for analyzing HER2 immunostaining, especially if they produce spurious or nonspecific results, or if they require frequent adjustments to enable the algorithm to calculate a HER2 score.

Validation is a formal system designed to gather and document evidence that provides a high degree of assurance that a process, system, or test method will consistently produce a result that meets predetermined acceptance criteria.⁷ Validation is intended to ensure that a product meets the operational needs of the user. According to the CAP Laboratory Accreditation Program (LAP) (ANP.23004 DIA preanalytic validation; ANP.22978 HER2 validation; TLC.10475 new instrument validation; COM.40300 accuracy validated; COM.40000 validation summary),⁸ the validation requirement applies to both new and existing assays. There are 2 major types of QIA systems: FDA-cleared/approved or laboratory-developed tests (LDTs), which carry different validation burdens (refer to Table 2). Validation must be performed on a minimum of 20 positive and 20 negative samples for FDA-cleared/approved assays that have been independently scored by several pathologists. For LDTs, validation must be performed on more than 40 positive and 40 negative samples. Equivocal samples (examined by both IHC and in situ hybridization [ISH] testing so that the true nature of the result is known) should also be included for completeness. Refer to Statement 2 below for further details. Principles for conducting a correct validation test have been published by Fitzgibbons et al^{9,10} and should be followed.

Regarding the scoring schema, one that matches the scoring system of the ASCO/CAP HER2 guideline (ie, score 0 and 1+ = negative; score 2+ = equivocal; score 3+ = positive) is preferable.

In the open comment period, there were 129 respondents, of whom 85.27% agreed (n = 110) and 14.73% (n = 19) disagreed with the draft guideline statement because they were confused about the specific nature of the validation requirement. This concern is further addressed in Statement 2.

2. Recommendation.—Laboratories should validate their QIA results for clinical use by comparing them to an alternative, validated method(s) such as HER2 fluorescence

in situ hybridization (FISH) or consensus images for HER2 IHC.

There were insufficient published data to inform this recommendation; however, current regulatory requirements, extensive clinical experience, and a strong expert consensus were deemed *adequate* to support this recommendation.

Validation of QIA for HER2 within a clinical laboratory is a requirement for both FDA-approved analytics and LDTs.¹¹ QIA tests must be validated by comparing results to an alternative method such as HER2 IHC images scored by an expert according to ASCO/CAP guidelines or FISH data. CAP checklist requirements that pertain to this recommendation include ANP.23004 DIA preanalytic validation; ANP.22978 HER2 validation; TLC.10475 new instrument validation; COM.40300 accuracy validated; and COM.40000 validation summary.⁸

The systematic literature review identified 1 study that addressed the issues of validation of QIA algorithms for HER2 IHC.¹² Bolton et al¹² assessed the agreement between automated IHC scores and pathologists' scores of negative versus positive and categorical strength of staining. To quantify the agreement between automated scores and pathologists' scores, these authors evaluated tissue microarrays from 440 breast cancers stained for HER2. The positive/negative class assignments between a consensus pathologist read and the output of 3 different algorithmic systems were compared by using the area under the receiver operating characteristics curve (ROC) and weighted κ statistics for categorical scores. The ROC curve plots the true-positive versus false-positive fraction for each possible cutoff point that could have been used to define negative versus positive tumors. To assess the agreement between the continuous scores of the automated instruments and the categorical scores of the pathologists for strength of staining, the authors converted the automated scores into the 4 categories used by the pathologists (negative, low, moderate, and strong staining). This was done by aligning the distributions of the pathologist semiquantitative scores with the automated scores. The agreement levels for HER2 were area under curve = .94 to .97 with κ = .53 to .72.¹²

Although the article authored by Al-Kofahi et al¹³ did not meet the inclusion criteria of our systematic review, this

particular study described an automated open-source cell-based system (FARSIGHT [www.farsight-toolkit.org]) and used a multispectral imaging camera (Nuance, CRL, Woburn, Massachusetts) to capture and analyze fluorochrome labels. This quantification of HER2 by QIA was validated in 2 ways. HER2 results were compared against determinations by a human expert and a validation was also made, based on cell cultures in vitro. For 3+ cases scored manually, using a revised threshold, the concordance rate was 97% and the positive predictive value was 97.6%.

Additionally, QIA may be applied to the analysis of the comparative methods used for HER2 IHC validation. van der Logt et al¹⁴ reported a validation study of QIA for HER2 FISH compared to manual FISH image analysis. The overall agreement with manual Abbott FISH data among tissue microarray samples and 50 selected IHC 2+ cases was 98.8% ($\kappa = .94$) and 93.8% ($\kappa = .88$), respectively.

Consequently, the combined reports demonstrate the breadth of different validation studies that can be applied to HER2 IHC validation. These may include the following:

1. Comparison with manual consensus scoring of IHC cases for HER2
2. Comparison with FISH numeric chromosome counts for *HER2*
3. Comparison with bright-field chromogenic in situ hybridization numeric chromosome counts for *HER2*
4. Comparison with a previously validated QIA algorithm for HER2.

The literature emphasizes the importance of performing such comparative studies in order to understand the differences in operating characteristics between manual and QIA methods.^{8,12-14}

In the open comment period, there were 110 respondents, of whom 93.64% ($n = 103$) agreed and 6.36% ($n = 7$) disagreed with the draft guideline statement. There were 26 written comments. While the survey responders overwhelmingly agree with this recommendation, their comments indicated varied opinions as to (1) what specifically was meant by alternative validated method, and (2) what are appropriate ground-truth alternative methods. Four of the responders indicated that FISH was the preferred alternative method for establishing ground truth; however, 11 of the responders favored the use of some IHC reference material (provided as standards that are scored by a consensus panel) as ground truth, and 11 of the responders were not clear as to what the "ground-truth" alternative method should be. One responder indicated that ground truth should be the product of both FISH and IHC reference materials. These comments were taken into consideration and the revisions resulting from these comments are reflected in the final wording of this guideline statement.

3. Recommendation.—Laboratories should ensure that the results produced by a QIA system are reproducible within and between different batch analyses.

There were insufficient published data to inform this recommendation; however, current regulatory requirements, extensive clinical experience, and a strong expert consensus were deemed *adequate* to support this recommendation.

Clinical Laboratory Improvement Amendments (CLIA) regulations stipulate that laboratories must establish and verify the performance specifications for all assays used in patient testing.¹⁵ An assessment of precision is required as a

part of this analytic validation process.¹⁵ Precision has been defined as "the closeness of agreement between independent results of measurements obtained under stipulated conditions."¹⁶ The CAP LAP specifies that precision be established by repeated measurement of samples or activities within-runs and between-runs over a period of time (All Common Checklist COM.40300).⁸ As applied to QIA, this would include an assessment of intrarun and interrater reproducibility.

While vendors of FDA-cleared or FDA-approved QIA systems may provide information on their instrument performance characteristics, it is still important to verify that these performance characteristics are reproducible in individual laboratories given the potential for variation in testing conditions, materials, and personnel.

While the specific methods used to determine QIA performance characteristics are ultimately at the discretion of individual laboratory directors, potential strategies are discussed below. Separate investigations should be conducted to assess system-induced variability as opposed to operator-induced variability associated with ROI selection (discussed in Statement 4). Case material selected for assessment of intrarun and interrater reproducibility should ideally include a "control" slide set derived from cases that are representative of each HER2 scoring category (ie, 0, 1+, 2+, 3+); such a slide set should also be representative of typical case material processed and seen in routine practice for a given laboratory. Serial formalin-fixed, paraffin-embedded HER2-stained sections of selected cases can be used to assess intrarun reproducibility, while interrater reproducibility can be assessed by scanning the same slide set on different days. For QIA systems that produce continuous instrument readings, the standard deviation and coefficient of variation can be calculated. Assessment of intrarun and interrater reproducibility should also be performed as a part of routine analytic validation of the QIA system, and needs to be verified after any significant changes to the standard operating procedure or instrumentation.

During the open comment period, we received 102 responses for this recommendation. Of those who responded, 95.1% ($n = 97$) indicated agreement with the draft recommendation, while 4.9% ($n = 5$) disagreed. Overall, there were 17 written comments with many responses expressing support for this particular recommendation. A number of responses suggested that additional detail be provided regarding the timing and specific requirements for assessment of within- and between-batch reproducibility. There were 5 comments suggesting that the QIA vendor should be responsible for providing documentation of within- and between-batch reproducibility. These comments were taken into consideration and while no changes were made to the guideline statement, these concerns are addressed in this article.

4. Recommendation.—Laboratories should ensure that the results produced by a QIA system are reproducible between operators when they select ROIs for analysis and/or perform annotation.

The strength of evidence was *adequate*.

Eight studies^{12,17-23} comprise the evidentiary base for this recommendation. The risk of bias assessment of the included studies ranged from low to high. None of the studies were found to have methodologic flaws that would call into question the study findings. Refer to Supplemental

Table 2 in the supplemental digital content for the quality assessment results.

Although QIA systems are designed to produce objective, accurate, and reproducible results, most of the instruments in current clinical use utilize algorithms that rely on operators to select ROIs to obtain a result. This process of ROI selection can introduce a source of interobserver variability. To overcome this problem, laboratories are encouraged to develop documented procedures for the training and selection of ROIs to be used in the validation study and for the concordance assessment (competency) of laboratory professionals or pathologists before testing clinical samples to ensure reproducibility of results in accordance with ASCO/CAP HER2 testing guidelines^{2,24} and accreditation requirements.⁸

Reproducibility, concordance, and observer agreement studies have shown moderate to high reproducibility in interobserver and intraobserver analysis using QIA systems for the scoring of HER2 by IHC.^{12,17–23} The study designs, HER2 testing modalities, QIA platforms, and algorithms in these publications vary considerably. The most relevant studies for this particular recommendation are those that specified the ROI selection process and that included clinical samples for analysis, therefore simulating clinical practice characteristics.^{19,20,22}

One QIA study used a validated method for the selection of ROI, using an FDA-approved manufacturer-locked algorithm (ie, no tuning possible) in 154 clinical samples, using 1 QIA instrument.¹⁹ The observers included surgical pathologists and cytotechnologists, but only the cytotechnologists performed manual as well as digital readings with the QIA system, and on more than 1 occasion in a subset of cases. They demonstrated that laboratory personnel were able to achieve high precision and accuracy with image analysis of HER2-stained slides, using their validated scoring technique (interobserver reproducibility κ score with cytotechnologist-assisted assessment, .77; intraobserver reproducibility κ score with cytotechnologist-assisted assessment, .88).¹⁹

Another study that used manufacturer-recommended methods for ROI selection and vendor-provided algorithms, and 2 QIA instruments (114 and 90 cases evaluated in each instrument, respectively) compared manual scoring with automated scoring for 2 pathologists, using each instrument on at least 2 occasions. These investigators demonstrated high interobserver and intraobserver reproducibility, with statistically significant intraobserver concordance for both instruments.²²

A multisite performance study of 260 breast tissue specimens compared a trainable IHC HER2 image analysis system with manual microscopy scored by 3 pathologists at 3 different sites.²⁰ The selection of ROI was described, but the authors acknowledged that a source of variation could have been the different hot spots chosen by each pathologist. Nevertheless, this study showed moderate to perfect agreement between different pathologists by manual readings (κ score, .48–.83) that was improved by image analysis readings with an interobserver near-perfect agreement (κ score, .73–.89).²⁰ Statistical significance was not reported for the κ values.

Laboratories testing for HER2 are required to demonstrate compliance with laboratory accreditation regulations that include ongoing competency assessment of laboratory professionals and/or pathologists interpreting a HER2 assay with a concordance of at least 95%. When using QIA

systems for HER2 testing this can be achieved by developing procedures that stipulate selecting ROIs and scoring. One consideration is that the scoring algorithms provided by a vendor may be locked owing to FDA clearance or approval, or may be amenable to tuning. For the latter algorithms, the standard operating procedures should address the extent of tuning allowed, if any, and laboratories should include this variable in their validation studies. The procedures for the ROI selection may follow vendor recommendations or be self-determined by each laboratory. In general, these methods should take into consideration the selection of areas of invasive carcinoma (or the intended neoplastic tissue when applicable), the exclusion of stroma or other nonepithelial cells depending on the algorithm, the inclusion of a representative number of neoplastic cells (absolute number or overall cellularity of the region selected), the selection of areas with representation of all staining patterns and intensities, as well as the incorporation of negative areas if no positive staining is identified. In one study for example, the validated scoring method included 6 regions that approximate the tissue that is included in an $\times 40$ objective, encompassing 2 areas of high-intensity staining, 2 areas of moderate-intensity staining, and 2 areas with low-intensity staining; when low-intensity staining areas were absent, 2 negative areas were used in substitution.¹⁹ In this study, the field diameter or cellularity of the regions was not specified. Other examples that support the above recommendation include the use of a free-form drawing tool to select a minimum of 6 areas of various sizes. The number of cells in each area was not specified.²²

Laboratories are responsible for the training of the laboratory professionals or pathologists on the ROI selection procedure. The documentation of procedure dissemination and training should follow CAP LAP or similar accreditation requirements.⁸ Once a procedure for the ROI selection has been developed, laboratories should validate the accuracy and precision of their procedure across a number of cases and operators before implementation.

In the open comment period, there were 99 respondents, of whom 89.9% ($n = 89$) agreed and 10.1% ($n = 10$) disagreed with the draft guideline statement. There were 17 written comments, including a number that emphasized that the major issue related to reproducibility was that the operators, not the image analysis systems, manually select the ROIs. Several respondents pointed out that the issue of manual selection of scoring areas as it relates to reproducibility is highly dependent on the standardization, training, and communication of this process. In addition, a number of comments stated that reproducibility should be imperative when using image analysis systems, that these QIA systems should not be inferior to manual scoring, and that analysis of the same regions should yield reproducible results. There was 1 comment stating that the manual tuning of algorithms should be considered when addressing reproducibility. These comments were taken into consideration, and, although no revisions were made into the final guideline statement, they are reflected in the discussion above.

5. Recommendation.—Laboratories should monitor and document the performance of their QIA system.

There were insufficient published data to inform this recommendation; however, current regulatory requirements, extensive clinical experience, and a strong expert consensus were deemed *adequate* to support this recommendation.

The CAP Anatomic Pathology Laboratory Checklist discusses the need for preanalytic testing phase validation (ANP.23004) as well as the use of appropriate slides for calibration for digital image analysis (ANP.23009).⁸ This checklist also requires daily quality control for control materials at more than 1 expression that are run concurrently with patient specimens (ANP.23018).⁸ Additionally, requirement ANP.23025 discusses the need for monthly quality control review of the control data by the laboratory director or designee.⁸ The CAP Laboratory General Checklist discusses the need for validation of the whole-slide imaging (WSI) system, specifically to be conducted by pathologists adequately trained to use the system (GEN.52920).⁸ This checklist also requires reevaluation of the entire WSI system if a significant change is made to the validated system.⁸

Although image analysis may be affected by various parameters including (but not limited to) tissue thickness, staining, and choice of whole slide scanner, the quantification results as assessed by visual inspection of the digital slide image should be in concordance to the results obtained by QIA. Laboratories should define an ongoing quality control process that monitors the results of HER2 attained by QIA imaging along with maintaining algorithm accuracy. Laboratories should also establish a quality control assurance program that evaluates the extent to which the image analysis results, when evaluated against a gold standard, are affected on account of changes to tissue processing, slide preparation, and/ or scanning.

In our review of the literature, one study²⁵ showed that the scoring of HER2 using the Automated Cellular Imaging System (ACIS, ChromaVision Medical Systems Inc, San Juan Capistrano, California) was associated with high false-positive rates before 2008. However, concordance rates for this institution improved subsequently, not only because they adopted newer QIA technology, but possibly also owing to greater care being taken with respect to tissue handling according to ASCO/CAP guidelines, with HER2 scoring via the image analysis system and assay standardization.

In a comparison study of image analysis for QIA of HER2 and estrogen receptor, multispectral-based methods yielded higher area-under-curve values compared to red, blue, green (RGB) images.²⁶ Additionally, the interobserver agreement and consistency assessment was more robust for the multispectral images than for the RGB images. These results appear to suggest that the image analysis system therefore needs to be optimized for the specific modality of image being analyzed.²⁶

Dennis et al²⁷ discussed the importance of instrument validation and calibration for achieving high concordance rates for quantification of HER2 using the VENTANA image analysis system (Roche, Basel, Switzerland). These authors found that there were substantial differences between the IHC results when using the manufacturer's machine score cutoffs versus laboratory-defined cutoffs with the FISH assay. The study highlighted the importance of instrument calibration to reduce false-positive results.²⁷

In the open comment period, there were 98 respondents, of whom 89.8% (n = 88) agreed and 10.2 % (n = 10) disagreed with the draft guideline statement. Some of the concerns raised by many of the participants included disagreement with the use of the word *continuous* with regard to monitoring of the laboratory QIA system. Other comments stressed the importance of closely monitoring

preimaging factors such as tissue section slice thickness and stain parameters, as opposed to just the QIA system parameters. While the expert panel agrees with these comments, this concept is outside the scope of this guideline. Further comments included integrating color standardization and resilience to slice tissue section thickness directly into the image analysis algorithms, thereby obviating or reducing the burden on the monitoring system. These comments were taken into consideration and the revisions are reflected in the final guideline statement presented in this document. The specific change instituted was to remove the term *continuous*, thereby removing the time dependency on the process. This would allow laboratory personnel to mitigate an unnecessary burden on staff for continuous monitoring and could allow them to invoke a need for monitoring or evaluation of the system in light of major changes to flow/practice.

6. Recommendation.—Laboratories should have procedures in place to address changes to the QIA system that could impact clinical results.

There were insufficient published data to inform this recommendation; however, current regulatory requirements, extensive clinical experience, and a strong expert consensus were deemed *adequate* to support this recommendation.

While the systematic literature review did not identify any studies that addressed documenting changes to a QIA system, proper change control procedures are well known concepts in the laboratory. The CAP All Common Checklist requires that all instruments and equipment be verified upon installation and after major maintenance or service (COM.30550).⁸ The CAP Laboratory General Checklist requirements GEN.43022 and GEN.43033 discuss changes to computer programs; specifically, they require documentation that programs are adequately tested for proper functioning when first installed and after any modifications, that the laboratory director or designee has approved the use of all new programs and modifications, that customized software, and modifications to that software, are appropriately documented, and that the records allow for tracking to identify persons that have added or modified any software.⁸

Laboratory procedures for change control will provide a formal process by which changes to the QIA system are introduced and controlled in a coordinated manner. This process will assure that any changes are documented and managed to prevent unintended consequences and that, ultimately, the image analysis algorithm maintains precision and accuracy. This procedure should encompass the description of the change, the reason for the change, the responsible individuals, the impact of the change, and the actions required. Table 3 describes the various types of changes and appropriate actions involved when using QIA. The change procedure should be governed by the medical director or a designee. Changes to the QIA system need to be evaluated, assessed, and reviewed for criticality. Revalidation is required when a major change (one that may impact the algorithm results) to the QIA system is implemented. Additionally, the expert panel believes that it is sound laboratory practice to treat changes of an uncertain impact level as a major change. Revalidation may consist of repeating the original validation effort or a portion of it, depending upon the nature and criticality of the change.

In the open comment period, there were 98 respondents, of whom 93.88% (n = 92) agreed and 6.12 % (n = 6)

Table 3. Potential Changes and Appropriate Actions for Quantitative Image Analysis Systems

Type of Change	Appropriate Action
Hardware replacement of "like for like" part	Document with equipment maintenance
Algorithm upgrade/version change	Documentation with revalidation
Hardware or software modification/upgrade to IHC platform	Documentation with revalidation
Hardware or software modification/upgrade to scanner	Documentation with revalidation
Modification/change to IHC protocol or 510(k)-cleared kit	Documentation with revalidation
Computer/server hardware or software upgrade/change	Documentation with revalidation or function check
Change to different IHC platform, antibody, scanner, or algorithm	Documentation with revalidation

Abbreviation: IHC, immunohistochemistry.

disagreed with the draft guideline statement. There were 12 written comments and 21 responses including a number that suggested that the guideline statement was too vague and that they did not understand what "change control" meant. Others commented that recommendations for revalidation requirements would be helpful, as would an explanation or definition of what comprises a major change, while others articulated that revalidation requirements should be the responsibility of the scanner and/or software manufacturer. These comments were taken into consideration, and, while no changes were made to the guideline statement, these concerns are addressed in this article.

7. Expert Consensus Opinion.—The pathologist should document that results were obtained by using QIA in the pathology report.

There were insufficient published data to inform this recommendation; however, regulatory requirements, extensive clinical experience, and a strong expert consensus were deemed *adequate* to support this guideline statement.

According to the CAP Anatomic Pathology Laboratory Checklist (ANP. 22969), the HER2 IHC report must include the type of specimen fixation and processing used, the antibody clone used in the detection system, the criteria used to determine a positive and negative result, and/or the scoring system that was used.⁸ This checklist also requires that the final report include the specimen sources, name of the vendor and imaging system used, the antibody clone or probe, and the detection method, as well as any limitations for the test result, if applicable (ANP.23038).⁸ It is the consensus of the expert panel that HER2 QIA results be reported by using the ASCO/CAP scoring schema² and that the report specify that QIA was used. Documenting that QIA was used will provide evidence for billing and quality monitoring purposes.

Laboratories may also wish to include in their report additional details such as a statement about adequacy (eg, "the digitized slide was deemed suitable for quantitative image analysis"), description and disclaimer of the methods (eg, if the method is FDA approved or represents an LDT, the cutoffs for a positive result), a comment or educational note, or an optional representative ROI image that was used in the analysis. The expert panel believed these items to be

too proscriptive to recommend and agreed that the individual laboratory policy indicate if such items should be included in the final pathology report. The expert panel did agree, however, that it is good laboratory practice to keep records of more detailed information about the image analysis system/algorithm, such as the version of the software used for testing. Items such as the version of the software are important for documentation purposes even if not included in the final report.

In the open comment period, there were 97 respondents, of whom 95.88% (n = 93) agreed and 4.12% (n = 4) disagreed with the draft guideline statement. There were 37 written comments, including 7 that suggested keeping the reporting requirements to a minimum. Others suggested inclusion of the image analysis platform and algorithm being used and their FDA approval status. These comments were taken into consideration and the revisions are reflected in the final guideline statement presented in this document.

8. Recommendation.—Personnel involved in the QIA process should be trained specifically in the use of the technology.

There were insufficient published data to inform this recommendation; however, current regulatory requirements, extensive clinical experience, and a strong expert consensus were deemed *adequate* to support this recommendation.

CAP General Checklist requirement GEN.55450 requires that records be maintained for all laboratory personnel, indicating that they have satisfactorily completed initial training on all instruments and methods applicable to their designated job.⁸ Additionally, requirement GEN.55500 requires that the competency of each person performing patient testing to perform his or her assigned duties be assessed.⁸

While there is currently little evidence of the impact of training in improving QIA, it is intuitive that this should be the case. However, there is sufficient evidence that, with appropriate training, pathologists and scientists can reliably perform, interpret, and report the results of HER2 ISH analyses.²⁸ A structured training about the QIA technology should ensure that personnel involved in QIA testing have consistent experience and background knowledge.

Personnel who are responsible for or involved in handling of an image analysis system and its data should be qualified as high-complexity testing personnel with experience as defined by the laboratory director. Records of qualifications including degree or transcript and work history in related fields should be maintained on file. Personnel involved in the QIA process should receive training enabling them to suitably perform the duties expected of an image analyst. This training could be conducted by structured training courses or by a qualified trainer who meets the recommendation of the laboratory director. Training from either the QIA vendor or a laboratory-developed training program should suffice. Records of training and maintenance of competency should be maintained on file.

In the open comment period, there were 97 respondents, of whom 92.78% (n = 90) agreed and 7.22% (n = 7) disagreed with the draft guideline statement. There were 9 written comments, most of which were unclear about what "training" means/entails. These comments were taken into consideration and have been described above.

9. Expert Consensus Opinion.—Laboratories should retain QIA results and the algorithm metadata in accordance with local requirements and applicable regulations.

Table 4. Classification of Metadata

Metadata Type	Definition	Mechanism Data Are Added	Examples
Technical metadata	Capture specific technical information pertaining to the image	Automatic	Device used, image size, color profile, device settings
Descriptive metadata	Information about the image	Manual	ROI, content labels, image notes, keywords, diagnosis, titles
Administrative metadata	Information pertaining to licensing, usage rights, and provenance	Manual	Creator, license holder, terms of use

Abbreviation: ROI, region of interest.

There were insufficient published data to inform this recommendation; the strength of evidence was *inadequate* to support specific recommendations on retention.

Quantitative image analysis performed on WSI files using a computational algorithm is a “laboratory test,” analogous to any other analytic test performed in the clinical laboratory. As such, the CLIA standard for laboratory regulations applies to the procedure, and laboratories that perform this procedure,¹⁵ and the retention requirements for glass slides used to prepare the digital image, the requisitions used to request the tests, the procedures describing the performance of the analysis, and the reports with the result or “interpretation” generated by the QIA test must adhere to the standards set forth by the CLIA laboratory standard.¹⁵

We were unable to find any specific evidence within the systematic literature review to support a specific recommendation regarding the retention of either the “regions of interest” or the entire WSI data used by the testing algorithm, or of the metadata generated by the algorithm (ie, the computation results) that is used to produce a result for the test. When deliberating a recommendation regarding storage needs related to QIA, it is important to consider the preanalytic, analytic, and postanalytic steps involved in the QIA process.

The preanalytic input for QIA is a digital image file, most often produced by a WSI capture device (scanner). Since 2010, the CAP’s LAP checklist for anatomic pathology has contained a section on digital image analysis, which primarily focuses on morphometric analysis, DNA analysis, and FISH.⁸ For the first time, the retention of images is regulated, although this is limited to images produced from slides that would otherwise be unreadable throughout the duration of the retention time requirement of 10 years (eg, FISH images).²⁴

When considering the analytic part of the QIA process, a short overview of metadata is required. Images are complex data objects; there is both a visual component and pixel binary data component that underly the visual representation. Image metadata is textual information that pertains to the image, and may either be “embedded” into the image file, or contained in a separate file associated with the image.²⁹ Image metadata can be further classified as technical, descriptive, or administrative (see Table 4). Typically, an image metadata standard format (eg, Exchangeable image file [Exif], the International Press Telecommunications Council [IPTC] Information Interchange Model [IIM], Dublin Core Metadata Initiative [DCMI], Extensible Metadata Platform [XMP]) is used to encode image metadata, allowing software applications to access and retrieve this information when necessary.

The analytic step of QIA usually begins with the designation of an ROI on which to apply the algorithm. The ROI is an example of descriptive metadata that designates a “subset” of the underlying pixel binary data. Commonly referred to as an annotation, this metadata is used by the image analysis algorithm as a “roadmap” for retrieving the pixel binary data of the image for input into the algorithm for computation. An algorithm will typically perform calculations on this input data, resulting in calculated output data. In reference to the original image, this calculated output data is also descriptive metadata that refers to the original image. For many algorithms, this output data is then mapped back to the original image by creating new “result” annotations (descriptive metadata) that are then visible in the software application. Refer to Figure 2 for the relationship of various data and metadata generated during the QIA process.

Once the algorithmic output has produced the results, the postanalytic step will be dependent upon the type of algorithm used. In some cases, the output itself may be used and reported. In certain instances, the output will be “interpreted,” either by a human or by computer, in conjunction with a standard/expected result to determine a “result” to be reported. In addition, metadata about the algorithm such as the version, name of the algorithm, the date the algorithm was applied, and the software vendor is required to be associated with the QIA process, as it provides context for interpreting the algorithm output results.

The aforementioned account of data and metadata creation involved throughout the QIA process shows the various parameters that should be considered in order to form a recommendation for data storage.

Given these parameters, a matrix (see Table 5) can be created detailing possible options to consider for data retention, each with pros and cons associated with the strategy regarding reproducibility, quality control, data reuse, storage needs, data management, future comparability, and clinical decision support.

After considering these potential options, and feedback from the community, the expert panel recommends retention of the algorithm’s computational output result data and the metadata about the algorithm used in accordance with local requirements and applicable regulations. The length of retention should be comparable to the current requirements for similar image assets, and based on documented standard operating procedures and policies. In the United States, the latest accreditation standard for data sets from *ex vivo* microscopic imaging systems is 10 years, and this is likely a good starting point

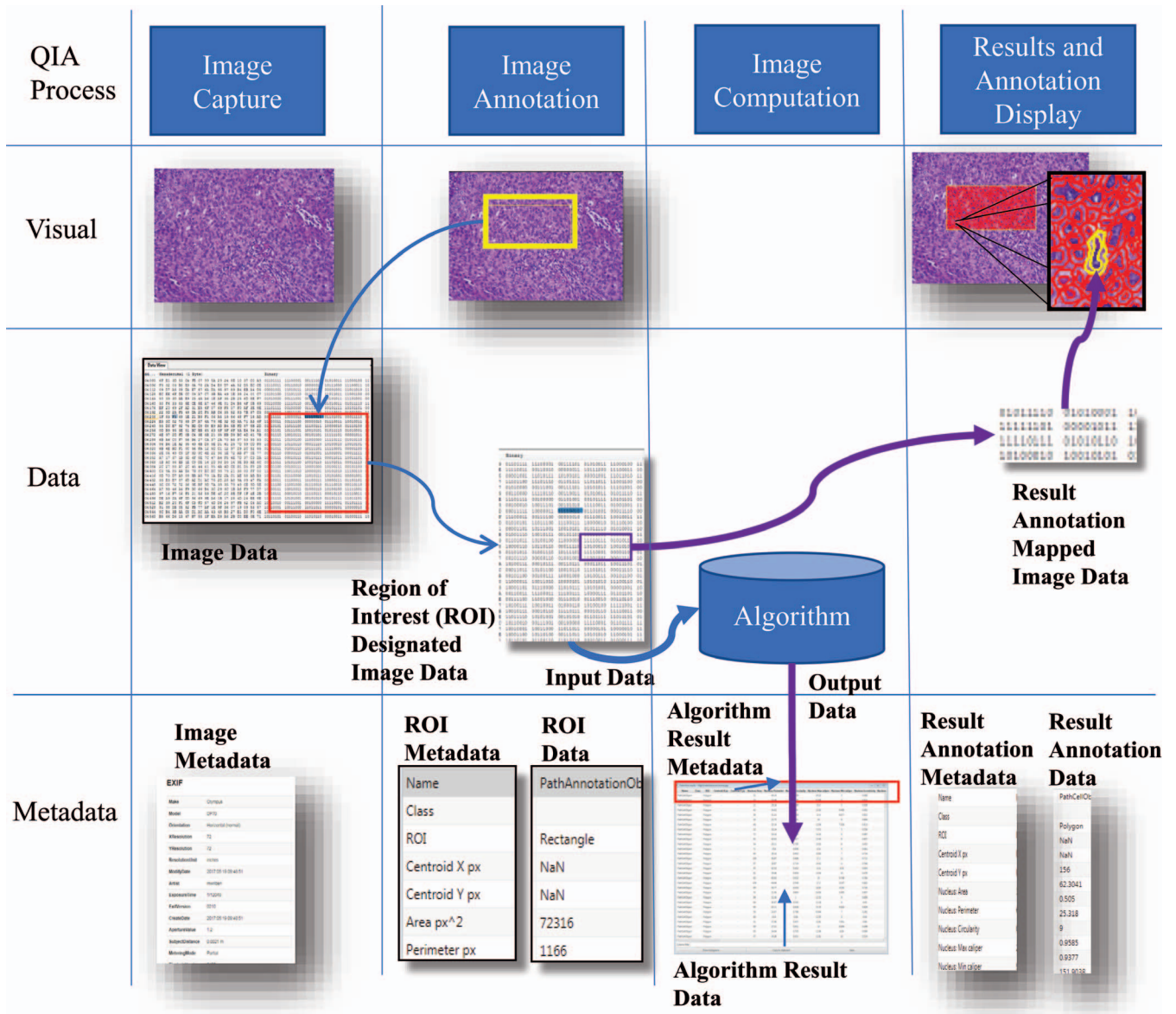


Figure 2. Relationship of various data and metadata generated during the quantitative image analysis (QIA) process. Abbreviation: ROI, region of interest.

for laboratories that do not have specific requirements at this time.⁸

In the open comment period, there were 95 respondents of whom 18 provided feedback regarding the draft guideline statement, which recommended retention of the WSI, the ROI metadata (annotations), the computational output results, and the metadata about the algorithm used; 83.16% (n = 79) agreed, while 16.84 (n = 16) disagreed with the draft guideline statement. There were 18 written comments received. While some simply agreed with the guideline statement, most comments expressed concern over the cost of QIA and the challenges related to storage of these data. There were also questions regarding responsibilities for laboratories that use third-party reference laboratories to perform this testing.

Several respondents were also concerned that the draft guideline statement imposed an undue burden on pathol-

ogists, given the cost and challenge of storing data above and beyond the cost required for retaining the glass slide and block, which is already mandated. The argument proposed that the “test” could be repeated if necessary by rescanning the slide or retrieving the original WSI. In the former case, it would be nearly impossible to replicate the exact ROIs designated by a human with a rescanned image. Even if the ROI selection is automated, it is likely impossible for the image data to be exactly the same as was used initially, particularly if a long period has occurred in between scans, given the nature of stain quality diminishing over time.³⁰

However, the expert panel recognizes that image data storage is a challenging and costly endeavor. Current commercially available systems lack functionality to reliably and easily retrieve ROIs and their associated metadata later within the context of the medical record. As a result, the

Table 5. Data Retention Strategy for Quantitative Image Analysis Assets					
	Data Retention Strategy				
	1	2	3	4	5
Option components					
Save glass slide	✓	✓	✓	✓	✓
Save WSI data file	✓	x	✓	x	✓
Save ROI metadata	x	✓	✓	x	✓
Save ROI raw data extracts	x	✓	✓	x	✓
Algorithm metadata	✓	✓	✓	✓	✓
Save algorithm output data	x	x	x	✓	✓
Pros					
Ability to repeat and generate the same results	x	✓	✓	x	✓
Apply new algorithms to original annotated areas	x	✓	✓	x	✓
Apply new or old algorithms to additional image annotation subsets	✓	x	✓	x	✓
Can compare new algorithm results to old easily	x	✓	✓	x	✓
Quick access to additional image subsets	✓	x	✓	x	✓
Saves data storage	x	✓	x	✓	x
Removes concern for future stain deterioration affecting image analysis	✓	x	✓	x	✓
Does not require advance software management capabilities	✓	x	x	x	x
Cons					
Never able to regenerate the exact same results	✓	x	x	✓	x
Cannot compare new algorithms to old ones easily	✓	x	x	✓	x
Potential for modified image data upon rescan due to stain deterioration	x	✓	x	✓	x
Does not allow quick access to other image subsets, and requires slide rescan	x	✓	x	✓	x
Requires more data storage	✓	x	✓	x	✓
Requires sophisticated software data management capabilities	x	✓	✓	✓	✓

Abbreviations: ROI, region of interest; WSI, whole slide imaging. Each column represents a possible data retention strategy. The components of each possible strategy are listed in the "Option components" section. ✓ = the option meets the property addressed; x = the option does not meet the property addressed.

expert panel revised our recommendation, decreasing the demands for pathologists to be in line with current system capabilities.

During the open comment period, issues with using cloud storage, cloud-based access to digital image analysis software, and obligations of the local laboratory that refers such QIA testing out to a third-party laboratory to perform were mentioned by some respondents. Best practices regarding cloud storage are beyond the scope of this guideline. However, this guideline statement makes no contradictory recommendations that would prohibit usage of storage in the cloud, or utilization of online QIA software systems.

10. Recommendation.—The pathologist who oversees the entire HER2 QIA process used for clinical practice should have appropriate expertise in this area.

There were insufficient published data to inform this recommendation; however, current regulatory requirements, extensive clinical experience, and a strong expert consensus were deemed *adequate* to support this recommendation.

According to the CAP Anatomic Pathology Laboratory Checklist, personnel responsible for evaluating or accepting the imaging system data must be qualified as high-complexity testing personnel (ANP.23041).⁸

As previously discussed, the personnel who oversee the HER2 QIA process may likely have a higher level of QIA expertise than a pathologist who signs out the HER2 QIA report. Nonetheless, the sign-out pathologist is still expected to recognize if the QIA system works as intended and if the results are valid for clinical purpose. The QIA expert should have the necessary skills and problem-solving abilities to address problems that may arise if the QIA system does not function as intended. This individual should also be able to supervise validation and monitor the QIA system. The latter may or may not be the same person who signs out the pathology report.

There are several complex steps involved in HER2 QIA. These include preimaging processes such as ensuring that there is good IHC staining, imaging tasks that involve image acquisition and calibration, performing of annotations, analysis (using software), and the generation of a meaningful pathology report to communicate the results of an image algorithm to the patient's treating clinician. Some image algorithms may require supervision for all of the steps in the process, whereas others may be automated from start to finish. There are many variables in this entire process (eg, the challenge of tissue heterogeneity), as well as sources of potential error (eg, artifacts), that can be introduced at any point.³¹ Hence, it is possible that an image analysis result may prove to be inaccurate when QIA is being used for clinical care. Hence, it is important that these vulnerabilities be addressed to avoid potential errors.

Accordingly, it is important that a qualified pathologist who is knowledgeable about all of the steps included in image analysis procedures be involved to oversee and critically assess the entire QIA process. This recommendation should not be interpreted to mean that only a pathologist is permitted to perform image analysis for clinical use. Rather, it is anticipated that a pathologist with expertise in image analysis will work closely with a qualified laboratory supervisor, proficient technologist, and/or other high-complexity testing personnel who are trained to execute 1 or all of the steps involved in performing QIA for clinical purposes. Although trained laboratory personnel may be responsible for evaluating and/or accepting imaging system data, the pathologist should oversee the quality management program established for his or her laboratory's QIA operation.

Similar to the requirements for a general laboratory medical director, this individual must have the necessary education, experience, and training.³² Specifically for QIA, they should be knowledgeable about all the steps involved including (1) the staining/immunohistochemical protocol used, (2) image analysis software functionality, (3) daily clinical operations and administration, (4) identification of potential errors, (5) indications of the final quantitative results, and (6) requirements for assuring compliance.

The published articles that were reviewed for this guideline unfortunately did not provide data that specifically addressed this recommendation for clinical practice.

In the open comment period, there were 94 respondents, of whom 74.47% (n = 70) agreed and 25.53 % (n = 24) disagreed with the draft guideline statement. Some written comments expressed concern about how best to train individuals to be proficient in QIA and that pathologist oversight of this process may increase workload (eg, having pathologists check each case).

11. Expert Consensus Opinion.—The pathologist finalizing the case should be knowledgeable in the use of the HER2 QIA system and visually verify that the correct ROI was analyzed, the algorithm-annotated image produced, and the image analysis results.

There were insufficient published data to inform this recommendation; however, extensive clinical experience and a strong expert consensus were deemed *adequate* to support this recommendation.

In contrast to the prior recommendation (Statement 10), which refers to the pathologist overseeing the entire QIA system at a practice, this recommendation applies to all the individual pathologists who are using the system to report out the results of a HER2 IHC test using QIA in clinical practice. While pathologists releasing or finalizing the QIA HER2 test results are not required to have advanced training in QIA, they should be familiar with the QIA system being used and are responsible for ensuring the quality of the test they are resulting using this system. They should also be qualified to interpret HER2 test results, which includes being familiar with the ASCO/CAP HER2 testing guideline,² HER2 IHC interpretation criteria, and being able to recognize unusual or discordant results. As such, this case-based quality assurance involves checking both the quality of the QIA on a given case (eg, ROI analyzed, the algorithm annotated image produced, and the image analysis results) and the IHC staining quality, plus correlation with the pathology findings of the case (according to ASCO/CAP HER2 testing guideline).² Recognition that there may be technical issues or unusual/discordant results may then result in additional consultation with specialists that have additional expertise in QIA, HER2 testing, or breast pathology, as appropriate to the presenting issue.

To ensure the quality of the preliminary QIA HER2 test result, there should be written criteria for acceptability of image analysis results and interpretation, and of the algorithm-annotated image produced. The image analysis results must be verified and reviewed by a qualified pathologist before reporting results. If the image analysis annotations do not yield acceptable results, the pathologist should have the oversight to troubleshoot the issue (including consultation with specialists) until there is resolution and not issue a report without further explanation. There should be a written policy in place stating that results must be reviewed and found to be acceptable before reporting patient results and stating the corrective actions to be taken if results are not acceptable.

The evidence identified by our systematic review included studies with reported practices resulting in high concordance rates of QIA HER2 IHC results with FISH test results.^{18,23,33–35} These studies provide indirect evidence that supports the importance of pathologist oversight of automated QIA results. Importantly, cases with QIA HER2 IHC interpretations that were discordant with FISH test results

were explored in some of these studies and highlight issues with QIA that require pathologist oversight and awareness. For example, cases called 3+ by QIA IHC analysis that had no gene amplification by *HER2* FISH were infrequent, but often had cytoplasmic or inconsistent staining patterns causing “false-positive” QIA IHC results.³⁴ Other common causes of such discordance (eg, QIA IHC 3+ but FISH negative) were attributed to unusual *HER2* FISH results such as cases with low-level increases in *HER2* copy number with coordinate increases in CEP17 control signals, which fall into a grey zone in *HER2* testing,³⁴ or borderline results near a threshold.¹⁸

In the open comment period, there were 94 respondents to the initial draft of this statement, of whom 78.72% (n = 74) agreed and 21.28% (n = 20) disagreed with the initial draft guideline statement. The original draft statement was worded such that a pathologist “trained in QIA must visually verify the image, the annotated image analysis output, and the algorithm results prior to finalizing the report.” There were 21 written comments that mainly focused on concerns about requiring additional formal training in QIA and confusion about how this would be defined. Therefore, the statement was revised to its current state to clarify that the pathologist finalizing the results of the QIA *HER2* test interpretation would not need formal training in QIA, but instead must at least be familiar with their particular QIA system and be able to verify that the image analysis was performed correctly and in the correct area (eg, ROI was selected appropriately for invasive cancer). Modifications to the initial statement were also made to make it clear that using a QIA system to help with *HER2* interpretation does not replace the ASCO/CAP *HER2* testing recommendation that the pathologist finalizing the *HER2* test should have met the competency assessment/performance requirements of his or her laboratory for interpreting *HER2* by IHC in breast cancers.² This pathologist should be able to recognize unusual *HER2* testing results/staining patterns (eg, heterogeneity, crush artifact, cytoplasmic rather than membranous staining) and also recognize if the results obtained are discordant with other findings in the case.

Some of the comments asked for more specific examples or standards of what criteria can be used to determine if a QIA result is acceptable. Each laboratory is responsible for creating a written policy of their criteria for acceptance/rejection of a test result and for performing and documenting ongoing competency assessment. However, suggestions common to the ASCO/CAP *HER2* testing guidelines² are emphasized:

1. Ensure that the area/ROI scored contains an appropriate area of invasive cancer for QIA analysis. This might include the highest-grade area of invasion with the strongest IHC expression.
2. Exclude regions containing crush or edge artifact, ductal carcinoma in situ, lymphocytes, or normal breast tissue from scoring.
3. Be aware of unusual staining patterns and histopathologic features suggestive of possible *HER2* test discordance (refer to Table 2 of 2013 ASCO/CAP *HER2* testing guideline).²

ADDITIONAL DISCUSSION

Although the expert panel was not tasked with addressing unusual *HER2* IHC staining and its impact in QIA, the panel

believed that mentioning such scenarios would help raise awareness of these issues and improve consistency in how the results are reported. As such, the following are examples of unusual HER2 IHC staining that the pathologist reporting the QIA test results should be aware of along with the panel's suggested guidance on how to resolve them:

1. Strong cytoplasmic staining only—consider calling equivocal (2+) and sending for ISH testing.
2. Heterogeneous staining (separate clustered areas with different staining patterns)—ensure that ROIs scored were in the strongest areas of staining. Report results based on correlation with the overall percentage of the invasive cancer with a similar staining pattern. If there is more than 10% 3+ staining, the case should be considered positive for HER2 overexpression. Pathologists should document the percentage that is positive and note that heterogeneity was present.
3. Discordant result with histology (eg, grade 1 or pure tubular or mucinous carcinoma that is HER2 3+)—consider reevaluating the grade/histologic subtype as well as the HER2 test. Additional testing on subsequent specimens may be required to resolve this discordant finding.

Additionally, a glossary of various key terms used throughout this article is provided in the supplemental digital content.

Limitations

The literature search did not address superiority or inferiority of QIA compared with manual interpretation, and considerations related to its utility such as cost and acceptability among laboratory personnel/administrators were not within the scope of this project. The literature search did not cover the engineering literature, where concepts such as the mechanics of optics and imaging systems might be addressed. Additionally, the existing pathology literature was relatively limited in the level of detail needed to answer the key questions used to formulate this guideline in greater detail. Preanalytic factors, while equally important, are not addressed in depth in this document. The focus of this guideline is narrow and does not address other testing methodologies such as ISH, FISH, immunofluorescence, and other biomarkers such as estrogen receptors, progesterone receptors, Ki-67, and p53. However, this process has provided valuable information to establish a framework and approach to investigate future guidelines on QIA of biomarkers of breast cancer as well as other cancers and diseases.

Future Considerations

There is a great deal of interest on the use of neural network and deep-learning-based technologies for QIA of digital pathology.³⁶ One of the possible concerns with these approaches is the apparent lack of transparency with respect to the features driving the QIA procedure. It is conceivable that these recommendations might need to be revisited in the future as deep-learning algorithms for QIA of digital pathology images continue to evolve.

CONCLUSIONS

To help laboratories implement QIA for clinical practice, we offer 11 guideline statements based on a systematic review of the literature. The major principles of the guideline

statements include the view that the image analysis system and procedures used must be validated before implementation and, once implemented, sustained with regular maintenance and ongoing quality control and quality assurance evaluation. Additionally, personnel involved in operating the QIA system should be trained in the technology and pathologists should supervise this process including the results.

As the literature continues to expand, this guideline will be reviewed and updated to ensure that the recommendations are relevant and sound, and to address new advances in the field.

Guideline Revision

This guideline will be reviewed every 4 years or earlier in the event of publication of substantive and high-quality evidence that could potentially alter the guideline recommendations. If necessary, the entire expert panel will reconvene to discuss potential changes. When appropriate, the expert panel will recommend revision of the guideline to the CAP for review and approval.

Disclaimer

The CAP developed the Pathology and Laboratory Quality Center as a forum to create and maintain evidence-based practice guidelines and consensus statements. Practice guidelines and consensus statements reflect the best available evidence and expert consensus supported in practice. They are intended to assist physicians and patients in clinical decision-making and to identify questions and settings for further research. With the rapid flow of scientific information, new evidence may emerge between the time a practice guideline or consensus statement is developed and when it is published or read. Guidelines and statements are not continually updated and may not reflect the most recent evidence. Guidelines and statements address only the topics specifically identified therein and are not applicable to other interventions, diseases, or stages of diseases. Furthermore, guidelines and statements cannot account for individual variation among patients and cannot be considered inclusive of all proper methods of care or exclusive of other treatments. It is the responsibility of the treating physician or other health care provider, relying on independent experience and knowledge, to determine the best course of treatment for the patient. Accordingly, adherence to any practice guideline or consensus statement is voluntary, with the ultimate determination regarding its application to be made by the physician in light of each patient's individual circumstances and preferences. CAP makes no warranty, express or implied, regarding guidelines and statements and specifically excludes any warranties of merchantability and fitness for a particular use or purpose. CAP assumes no responsibility for any injury or damage to persons or property arising out of or related to any use of this statement or for any errors or omissions.

We thank advisory panel members David Rimm, MD, PhD, Kenneth J. Bloom, MD, Richard Levenson, MD, Stephen Hewitt, MD, PhD, and Mogens Vyberg, MD, for their thoughtful feedback and review of this work. We also thank Patrick Fitzgibbons, MD, for his guidance throughout the project.

References

1. Breastcancer.org. U.S. Breast Cancer Statistics. Breastcancer.org Web site. http://www.breastcancer.org/symptoms/understand_bc/statistics. Updated January 9, 2018. Accessed July 13, 2018.

2. Wolff AC, Hammond ME, Allison KH, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/ College of American Pathologists clinical practice guideline focused update. *J Clin Oncol*. 2018;36(20):2105–2122.
3. College of American Pathologists. *Surveys and Anatomic Pathology Education Programs: Histotechnology Quality Improvement Program HQIP-A 2016*. Northfield, IL: College of American Pathologists; 2016.
4. Combrinck M, Fine J, Pantanowitz L. Comparison of Companion Virtuoso and stand-alone Visiopharm breast marker image analysis software [Pathology Informatics 2015]. *J Pathol Inform*. 2015;6:S23–S24.
5. Pantanowitz L, Liu C, Huang Y, Guo H, Rohde GK. Impact of altering various image parameters on human epidermal growth factor receptor 2 image analysis data quality. *J Pathol Inform*. 2017;8:39.
6. Institute of Medicine (IOM). *Clinical Practice Guidelines We Can Trust*. Washington, DC: The National Academies Press; 2011.
7. The Global Harmonization Task Force Study Group 3. *Quality Management Systems – Process Validation Guidance*. 2nd ed. [http://www.imdrf.org/docs/ghtf/final/sg3/technical-docs/ghtf-sg3-n99-10-2004-qms-process-guidance-04010.pdf#search="process validation"](http://www.imdrf.org/docs/ghtf/final/sg3/technical-docs/ghtf-sg3-n99-10-2004-qms-process-guidance-04010.pdf#search=). Published January 2004. Accessed July 16, 2018.
8. College of American Pathologists. Accreditation checklists. College of American Pathologists Web site. <https://www.cap.org/laboratory-improvement/accreditation/accreditation-checklists>. Accessed July 16, 2018.
9. Fitzgibbons PL, Bradley LA, Fatheree LA, et al. Principles of analytic validation of immunohistochemical assays: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med*. 2014;138(11):1432–1443.
10. Fitzgibbons PL, Murphy DA, Hammond ME, Allred DC, Valenstein PN. Recommendations for validating estrogen and progesterone receptor immunohistochemistry assays. *Arch Pathol Lab Med*. 2010;134(6):930–935.
11. US Food and Drug Administration. *Analytical Procedures and Methods Validation for Drugs and Biologics: Guidance for Industry*. Silver Spring, MD: US Food and Drug Administration, US Department of Health and Human Services; 2015.
12. Bolton KL, Garcia-Closas M, Pfeiffer RM, et al. Assessment of automated image analysis of breast cancer tissue microarrays for epidemiologic studies. *Cancer Epidemiol Biomarkers Prev*. 2010;19(4):992–999.
13. Al-Kofahi Y, Lassoued W, Grama K, et al. Cell-based quantification of molecular biomarkers in histopathology specimens. *Histopathology*. 2011;59(1):40–54.
14. van der Logt EM, Kuperus DA, van Setten JW, et al. Fully automated fluorescent in situ hybridization (FISH) staining and digital analysis of HER2 in breast cancer: a validation study. *PLoS One*. 2015;10(4):e0123201.
15. Department of Health and Human Services, Centers for Medicare and Medicaid Services. Clinical Laboratory Improvement Amendments of 1988; final rule. Standard: establishment and verification of performance specifications. *Fed Regist*. 2003;68(16):3707. Codified at 42 CFR §493.1253.
16. Linnet K, Boyd JC. Selection and analytical evaluation of methods – with statistical techniques. In: Burtis CA, Bruns DE, eds. *Tietz Fundamentals of Clinical Chemistry and Molecular Diagnostics*. 7th ed. St Louis, MO: Elsevier Saunders; 2015:6–32.
17. Gavrielides MA, Gallas BD, Lenz P, Badano A, Hewitt SM. Observer variability in the interpretation of HER2/neu immunohistochemical expression with unaided and computer-aided digital microscopy. *Arch Pathol Lab Med*. 2011;135(2):233–242.
18. Gustavson MD, Bourke-Martin B, Reilly D, et al. Standardization of HER2 immunohistochemistry in breast cancer by automated quantitative analysis. *Arch Pathol Lab Med*. 2009;133(9):1413–1419.
19. Minot DM, Voss J, Rademacher S, et al. Image analysis of HER2 immunohistochemical staining: reproducibility and concordance with fluorescence in situ hybridization of a laboratory-validated scoring technique. *Am J Clin Pathol*. 2012;137(2):270–276.
20. Nassar A, Cohen C, Agersborg SS, et al. Trainable immunohistochemical HER2/neu image analysis: a multisite performance study using 260 breast tissue specimens. *Arch Pathol Lab Med*. 2011;135(7):896–902.
21. Prasad K, Tiwari A, Ilanthodi S, Prabhu G, Pai M. Automation of immunohistochemical evaluation in breast cancer using image analysis. *World J Clin Oncol*. 2011;2(4):187–194.
22. Slodkowska J, Filas V, Buszkiewicz E, et al. Study on breast carcinoma Her2/neu and hormonal receptors status assessed by automated images analysis systems: ACIS III (Dako) and ScanScope (Aperio). *Folia Histochem Cytobiol*. 2010;48(1):19–25.
23. Turashvili G, Leung S, Turbin D, et al. Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent in situ hybridization (FISH): pathologist assessment compared to quantitative image analysis. *BMC Cancer*. 2009;9:165.
24. Lange HPD. Digital pathology: a regulatory overview. *Lab Med*. 2011;42(10):587–591.
25. Sarode VR, Xiang QD, Christie A, et al. Evaluation of HER2/neu status by immunohistochemistry using computer-based image analysis and correlation with gene amplification by fluorescence in situ hybridization assay: a 10-year experience and impact of test standardization on concordance rate. *Arch Pathol Lab Med*. 2015;139(7):922–928.
26. Liu WL, Wang LW, Chen JM, et al. Application of multispectral imaging in quantitative immunohistochemistry study of breast cancer: a comparative study. *Tumour Biol*. 2016;37(4):5013–5024.
27. Dennis J, Parsa R, Chau D, et al. Quantification of human epidermal growth factor receptor 2 immunohistochemistry using the Ventana Image Analysis System: correlation with gene amplification by fluorescence in situ hybridization: the importance of instrument validation for achieving high (>95%) concordance rate. *Am J Surg Pathol*. 2015;39(5):624–631.
28. Bartlett JMS, Starczynski J, Atkey N, et al. HER2 testing in the UK: recommendations for breast and gastric in-situ hybridisation methods. *J Clin Pathol*. 2011;64(8):649–653.
29. Rouse M. Definition: image metadata. Whatsl.com Web site. <http://whatis.techtarget.com/definition/image-metadata>. Updated June 2015. Accessed July 16, 2018.
30. Hanna M, Duboy J, Pantanowitz L. Whole slide image file integrity: a 10-year look back at archival images [Pathology Informatics Summit 2017]. *J Pathol Inform*. 2017;8(26):S48–S49.
31. Conway CM, Hewitt SM. Image analysis applications. In: Pantanowitz L, Parwani AV, eds. *Digital Pathology*. Chicago, IL: ASCP Press; 2017:153–172.
32. Kothari S, Phan JH, Stokes TH, Wang MD. Pathology imaging informatics for quantitative analysis of whole-slide images. *J Am Med Inform Assoc*. 2013;20(6):1099–1108.
33. Ayad E, Mansy M, Elwi D, Salem M, Salama M, Kayser K. Comparative study between quantitative digital image analysis and fluorescence in situ hybridization of breast cancer equivocal human epidermal growth factor receptors 2 score 2(+) cases. *J Pathol Inform*. 2015;6:31.
34. Holten-Rossing H, Moller Talman ML, Kristensson M, Vainer B. Optimizing HER2 assessment in breast cancer: application of automated image analysis. *Breast Cancer Res Treat*. 2015;152(2):367–375.
35. Tawfik O, Kimler BF, Davis M, et al. Grading invasive ductal carcinoma of the breast: advantages of using automated proliferation index instead of mitotic count. *Virchows Arch*. 2007;450(6):627–636.
36. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. *Med Image Analysis*. 2016;33:170–175.

APPENDIX 1. Disclosed Interests and Activities From October 2014 to November 2017

Name	Interest/Activity Type	Entity
Marilyn Bui, MD, PhD	Board, advisory boards	Digital Pathology Association (DPA) <i>Pathology and Laboratory Medicine International</i> <i>International Journal of Clinical and Experimental Pathology</i> <i>Journal of Pathology Informatics</i> <i>Sarcoma</i> <i>Open Journal of Pediatrics</i> <i>OA Case Reports</i>
	Consultancies	Rosen Louik & Perry, PC de la Parte & Gilbert, PA
	Grants/collaborative agreements	National Institutes of Health (NIH)/National Cancer Institute
	Leadership in other associations	DPA
	Elected or appointed positions in other national/international medical organizations	Chinese American Pathologists Association Florida Society of Pathologists American Society of Cytopathology (ASC)
	Lecture fees (honorarium)	Pathology Review, Osler Institute ASC
		American Society of Clinical Pathology
		Beijing Xiehe Hospital in Beijing
		Shandong University Qilu Hospital
		Chinese Orthopaedic Association
	Intellectual properties/patents	PCT patent application No. PCT/US2011/048995 US patent application No. 13/813,605 US patent application No. 61/588,215 US patent application No. 61/733,502 PCT application No. PCT/US2013/022017 US patent application No. 14/373,277 US patent application No. 14/373,277
	Royalties	Pathology Review, Osler Institute CAP Press, donated to CAP Foundation Demos Medical American Society for Clinical Pathology
Elizabeth Chlipala, BS, HTL(ASCP)QIHC	Board, advisory boards	Leica Biosystems
	Lecture fees (honorarium)	Leica Biosystems DPA
	Elected or appointed positions in other national/international medical organizations	National Society for Histotechnology Colorado BioScience Association
M. Elizabeth Hammond, MD Anant Madabhushi, PhD	Board, advisory boards	<i>Archives of Pathology & Laboratory Medicine</i>
	Board, advisory boards	Inspirata, Inc.
		Elucid Bioimaging
		Ibris, Inc
		AstraZeneca
	Consultancies	Diagnostic Precision Inc.
		Inspirata, Inc.
		Ibris, Inc
	Lecture fees (honorarium)	The Mount Sinai Hospital, New York, NY
		University of Cincinnati
	Grants/collaborative agreements	Siemens Philips General Electric
	Intellectual properties/patents	Elucid Bioimaging PCT/US12/20636 USSN: 9,002,092 Elucid Bioimaging PCT/US12/21373 Case No. 2013-2471 Elucid Bioimaging US 2013/0202173 Elucid Bioimaging USSN: 8,204,315 Elucid Bioimaging USSN: 8,774,479
Ownership or beneficial ownership of stocks	Inspirata, Inc Elucid Bioimaging	

APPENDIX 1. Continued

Name	Interest/Activity Type	Entity	
Liron Pantanowitz, MD	Board, advisory boards	Cambridge Healthtech Institute Leica Biosystems	
	Consultancies	Cambridge Healthtech Institute Vanderbilt University Hologic, Inc Hamamatsu Photonics K.K. Ibex Medical Analytics	
	Leadership in other associations	Association for Pathology Informatics (API) ASC	
	Lecture fees (honorarium)	ASC American Society for Clinical Pathology (ASCP)	
	Royalties	ASCP Press Springer UpToDate API	
	Michael W. Riben, MD	Leadership in other associations	API
	Mohamed E. Salama, MD	Board, advisory boards	Leica Biosystems Incite Inc.
Consultancies		Leica Biosystems Roche Diagnostics Seattle Genetics	
Lecture fees (honorarium)		Seattle Genetics	
Ownership or beneficial ownership of stocks		Techcyte	
John E. Tomaszewski, MD		Board, advisory boards	Ibris, Inc Inspirata, Inc.
		Leadership in other associations	Association of Pathology Chairs Intersociety Council for Pathology Information
		Grants/collaborative agreements	NIH
	Lecture fees (honorarium)	American Society for Investigative Pathology	
	Intellectual properties/patents	Imaging Network of Ontario	
	Ownership or beneficial ownership of stocks	University of Pennsylvania – 4 patents Inspirata, Inc	

The following individuals have reported no relevant disclosures: Kimberly H. Allison, MD; Andrea G. Kahn, MD; Rachel L. Stewart, DO, PhD; Carol Colasacco, MLIS, SCT(ASCP); Christina Lacchetti, MHSc; Nicole E. Thomas, MPH, CT(ASCP)^{cm}.