COLLEGE of AMERICAN PATHOLOGISTS
Laboratory Quality Solutions

Supplemental Digital Content* | Methodology | January 2019

Quantitative Image Analysis of HER2 Immunohistochemistry for Breast Cancer

Authors:
Carol Colasacco, MLIS, SCT(ASCP)
Marilyn Bui, MD, PhD
Christina Lacchetti, MHSc
Nicole E. Thomas, MPH, CT(ASCP)cm

Bui MM, Riben MW, Allison KH, et al. Quantitative image analysis of human epidermal growth factor receptor 2 immunohistochemistry for breast cancer: guideline from the College of American Pathologists. Arch Pathol Lab Med. 2019;143(10):1180-1195.
doi: 10.5858/arpa.2018-0378-CP

*The Supplemental Digital Content was not copyedited by Archives of Pathology & Laboratory Medicine.

METHODS USED TO PRODUCE THE GUIDELINE

Panel Composition

The College of American Pathologists (CAP) convened an expert panel (EP) consisting of pathologists, histotechnologists, and computer scientists with expertise in digital pathology (specifically image analysis), breast pathology, immunohistochemistry (IHC), and quality management, and a methodologist consultant to develop an evidence-based guideline to address quantitative image analysis for human epidermal growth factor receptor 2 (HER2) immunohistochemistry for breast cancer. CAP approved the appointment of the project chair and panel members. The EP members performed the systematic evidence review, drafted the recommendations, evaluated the public comments, revised the recommendations and contributed to the manuscripts.

An advisory panel (AP) of five pathologists also helped in the development of the guideline. The role of the AP members was to provide guidance and feedback on the scope and key questions for the literature search, vet the draft guideline statements prior to the public comment period, and to review and provide feedback for the manuscript and supplemental digital content (SDC).

Conflict of Interest (COI) Policy

Prior to acceptance on the expert or advisory panel, potential members completed the CAP conflict of interest (COI) disclosure process, whose policy and form (in effect April 2010) require disclosure of material financial interest in, or potential for benefit of significant value from, the guideline's development or its recommendations 12 months prior through the time of publication. The potential members completed the COI disclosure form, listing any relationship that could be interpreted as constituting an actual, potential, or apparent conflict. The CAP Center uses the following criteria:

Nominees who have the following conflicts may be excused from the panel:

a. Stock or equity interest in a commercial entity that would likely be affected by the guideline or white paper
b. Royalties or licensing fees from products that would likely be affected by the guideline or white paper
c. Employee of a commercial entity that would likely be affected by the guideline or white paper

Nominees who have the following potentially manageable direct conflicts may be appointed to the panel:

a. Patents for products covered by the guideline or white paper
b. Member of an advisory board of a commercial entity that would be affected by the guideline or white paper
c. Payments to cover costs of clinical trials, including travel expenses associated directly with the trial
d. Reimbursement from commercial entity for travel to scientific or educational meetings

Everyone was required to disclose conflicts prior to beginning and continuously throughout the project's timeline. The CAP provided funding for the administration of the project; no industry funds were used in the development of the guideline.

Disclosures of interest judged by the oversight group to be conflicts are as follows: AM, stock options/bonds, consultant/advisory fees, and research grants, Inspirata, Inc. (Tampa, FL); stock options/bonds, Elucid Bioimaging, (Wenham, MA); research grants, Philips Healthcare (Koninklijke Philips [Amsterdam, The Netherlands]); grants, Pathcore Inc. (Toronto, ON, Canada); JT, stock options/bonds, Inspirata, Inc. (Tampa, FL); LP, consulting or advisory fees,

Hamamatsu Photonics K.K., (Hamamatsu City, Japan); MS, stock options/bonds, Techcyte, Inc. (Orem, UT). The majority of the EP (seven of 11 members) was assessed as having no relevant conflicts of interest.

Systematic Evidence Review (SER)

The objective of the SER was to identify articles of sufficient quality that would provide data to inform the recommendations. The scope of the SER and the key questions (KQs) were established by the EP in consultation with the methodologist prior to beginning the literature search.

Search and Selection

An initial systematic literature search for relevant evidence was completed on 3/18/16 utilizing Ovid MEDLINE (Ovid Technologies Inc, New York City, NY). The search strategy included medical subject headings (MeSH) and text words to capture the general concepts of "quantitative image analysis", "HER2 testing", and "immunohistochemistry." Limits were set for the publication dates 1/1/2006 through 3/18/16 for human studies published in English. Commentaries, letters, and editorials were excluded from the original literature search. Supplementary searches were completed on 3/21/16 utilizing PubMed (National Library of Medicine, Bethesda, MD) and Scopus (Elsevier, B.V., Amsterdam, The Netherlands) for articles published from 1/1/2006 through 3/21/16. The PubMed and Scopus searches were adapted from the Ovid search string, and all search strings are included in Supplemental Figure 2.

Database searches were supplemented by additional searches for indexed and unindexed (grey) literature. The Cochrane Library, Google Scholar, the TRIP database, and the websites of guideline repository sites (https://guidelines.gov and www.g-i-n.net) were searched for relevant articles or guidelines. In addition, a focused search of organizations' websites (Clinical & Laboratory Standards Institute [CLSI], College of American Pathologists [CAP], American Society for Clinical Pathology [ASCP], Association of Directors of Anatomic and Surgical Pathology [ADASP], Association for Molecular Pathology [AMP], Digital Pathology Association [DPA], Royal College of Pathologists [RCP], Canadian Association of Pathologists, American Telemedicine Association, Digital Imaging and Communications in Medicine [DICOM], American Society of Clinical Oncology [ASCO], United States Food and Drug Administration [FDA], International Organization for Standardization [ISO]) was completed to identify applicable guidelines, protocols, or standards. Expert panel members were polled for information regarding any known unpublished studies of interest or additional published studies or guidelines not already identified. Search limits for all manual searches included human studies published in English from 1/1/2006 through 3/31/2016. All publication types except for letters, editorials, and commentaries were included in the literature search.

The Ovid MEDLINE search was rerun on 01/27/2017 to identify articles published from 3/18/16 through 01/27/2017 and guidelines repository sites (https://guidelines.gov and www.g-i-n.net) were reviewed on 01/27/2017 for newly published guidelines. A review of the reference lists of all articles included for data extraction was completed in order to identify relevant studies not previously identified.

All unique articles identified were added to DistillerSR [Evidence Partners, Ottawa, Canada] for review.

Selection at all levels was based on predetermined inclusion/exclusion criteria.

Eligible Study Designs

Practice guidelines, consensus documents, systematic reviews, meta-analyses, randomized control trials, comparative studies, reviews, and evaluation studies were eligible for inclusion. Meeting abstracts, non-comparative studies, and commentaries, editorials and letters were excluded a priori.

Inclusion Criteria

Published studies were included if they met each of the following criteria:

- Studies that address quantitative image analysis
- Studies that are focused on surgical pathology samples from the breast
- Studies that address HER2 testing using immunohistochemistry
- Comparative studies or guidelines, protocols or standards
- Studies that report on at least one outcome of interest
- Studies with a population of at least 20 patients/specimens/samples
- Clinical studies
- Human studies
- Studies published in English
- Studies published since 2006

Exclusion Criteria

- Animal studies
- Non-English studies
- Studies that address pathology samples from other body sites than breast
- Studies that address Immunofluorescence or testing/staining methods other than immunohistochemistry
- Studies with fewer than 20 patients/specimens/samples
- Research-based studies
- Studies published prior to 2006
- Studies that do not address HER2 testing
- Non-comparative studies other than relevant protocols, standards, guidelines
- Studies that do not address at least one outcome of interest
- Studies that do not address quantitative image analysis
- Meeting abstracts

Outcomes of Interest

The EP deemed the following as outcomes of interest: concordance rates/observer agreement (intra- or inter-observer agreement), scoring of HER2 IHC, reproducibility, accuracy (sensitivity and or specificity rates). Qualitative or quantitative data on the following items were also deemed as outcomes of interest: validation/verification/set-up, equipment function and calibration, controls such as illumination, throughput monitoring, selecting an algorithm, machine-specific issues, experience of the observers, training or competency assessments, storage of image or data, reporting, adequacy of image or image display information (eg, pixels, screen diagonal), staining intensity, and clinical outcomes.

Data Extraction & Management

The data elements from an included article/document were extracted by one reviewer into standard data formats and tables developed using the systematic review database software, DistillerSR (Evidence Partners Inc., Ottawa, Canada); a second reviewer confirmed accuracy and completeness. Any discrepancies in data extraction were resolved by discussion between the chair and the methodologist. A bibliographic database was established in EndNote (Thomson Reuters, Carlsbad, CA) to track all literature identified and reviewed during the study.

Literature Review and Analysis

The EP met 21 times through teleconference webinars from August 4, 2015, through November 29, 2017. Additional work was completed via electronic mail. The panel met in person January 14, 2017, to review evidence from the systematic review and to draft recommendations and again June 24, 2017, to revise the draft manuscript.

The EP formed the following key questions (KQs) for which to base the literature search:

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

Seven of the 11 EP members participated in the systematic evidence review (SER): title-abstract screening, full-text review, and data extraction of high-level studies (i.e., randomized control trials, systematic reviews, and clinical practice guidelines). A dual review was performed for each study and in each phase of the SER; another EP member (usually the chair) adjudicated all conflicts. Five EP members participated in the literature refresh, where the studies also underwent duel review. A total of 391 studies comprised the final body of studies included in the SER. Supplemental Figure 1 displays the results of the literature review. All articles were available as discussion or background references. All members of the EP participated in developing draft recommendations, reviewing open comment feedback, finalizing and approving the final recommendations, and writing/editing of the manuscript.

Peer Review

A public, open access comment period was held from March 6 through March 27, 2017 on the CAP Web site www.cap.org for any interested stakeholder to provide feedback on the draft recommendations. Eleven draft recommendations (along with topics that were to be covered in the manuscript), two demographic questions, three questions to assess feasibility, and one area to capture general comments were posted for feedback. An announcement was sent to the following societies deemed to have interest:

Medical Societies

- American Association of Pathologists' Assistants (AAPA)
- American Cancer Society (ACS)
- Association of Directors of Anatomic and Surgical Pathology (ADASP)
- Association for Molecular Pathology (AMP)
- Association of Pathology Chairs (APC)
- Association for Pathology Informatics (API)
- American Society of Cytopathology (ASC)
- American Society of Clinical Oncology (ASCO)
- American Society for Cytotechnology (ASCT)
- American Society for Clinical Pathology (ASCP)
- California Society of Pathologists (CSP)
- College of American Pathologists (CAP)
- Canadian Association of Pathologists (CAP-APC)
- Chinese American Pathologists Association (CAPA)
- Clinical Laboratory Management Association (CLMA)
- Digital Pathology Association (DPA)
- European Society of Pathology (ESP)
- Florida Society of Pathologists (FSP)
- International Society of Breast Pathology (ISBP)
- National Society for Histotechnology (NSH)
- New York Pathological Society (NYPS)
- Society to Improve Diagnosis in Medicine (SIDM)
- The Royal College of Pathologists (RCPath)
- Texas Society of Pathologists (TSP)
- United States and Canadian Academy of Pathology (USCAP)
- United Kingdom National External Quality Assessment Service (UK NEQAS)

- Nordic Immunohistochemical Quality Control (NordiQC)
- Canadian Immunohistochemistry Quality Control (CIQC)
- Clinical Laboratory Standards Institute (CLSI)
- IHC World

Other Societies/Organizations

- American Society for Investigative Pathology (ASIP)
- American College of Veterinary Pathologists (ACVP)
- Federation of American Societies for Experimental Biology (FASEB)
- Veterinary Cancer Society (VCS)
- European College of Veterinary Clinical Pathology (ECVCP)/European Society of Veterinary Clinical Pathology (ESVCP)
- European College of Veterinary Pathologists (ECVP)

Government

- US Food and Drug Administration (FDA)
- Centers for Medicare & Medicaid Services (CMS)
- U.S. Department of Veteran's Affairs (VA) and U.S. Department of Defense (DoD)

"Agree" and "Disagree" responses were captured for every proposed recommendation. The website also received 186 written comments. Five draft recommendations achieved more than 90% agreement, four draft statements achieved more than 80% agreement, and two received more than 70% achievement. Each EP member was assigned one draft statement for which they reviewed the comments and present them to the entire panel for group discussion. After consideration of the comments, five draft recommendations were maintained with the original language and six were revised. Resolution of all changes was obtained by majority consensus of the panel using nominal group technique (discussion at an in-person meeting, rounds of teleconference webinars, email discussion and multiple edited recommendations) amongst the panel members. The final recommendations were agreed upon by the EP with a formal vote. The panel considered laboratory efficiency and feasibility throughout the entire considered judgment process. 53.41% (47 of 88) responded that all of the draft guideline was feasible, 40.91% (36 of 88) responded that parts of it were feasible, and 5.68% (5 of 88) responded that none of it was feasible. Neither formal cost analysis nor cost effectiveness models were performed.

An independent review panel (IRP) was assembled to review and approve the guideline on behalf of the CAP Council on Scientific Affairs. The IRP was masked to the EP and to each other and were vetted through the COI process.

Quality Assessment Methods

An assessment of study quality was performed for all fully published studies meeting inclusion criteria by a research methodologist. Formal quality assessment involved determining the risk of bias by assessing key indicators, based on study design and methodological rigor. These items were assessed as being either yes (√), no (X), or not reported (NR). Methodological criteria assessed were informed by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) 2 tool,[1] with the following items considered:

- Sampling method used
- Blinding reported
- Same reference standard applied to all study subjects
- Appropriate statistical analyses reported

- Funding source

Each study was assessed individually and a summary of the overall quality of the evidence was given considering the evidence in totality.

A rating for the strength of evidence is given for guideline statements where quality was assessed (i.e., only studies obtained from our SR). Ultimately, the designation (rating) of the strength of evidence is a judgment by the expert panel of their level of confidence that the evidence from the studies informing the recommendations reflects true effect. Supplemental Table 1 describes the grades for strength of evidence.

Quality Assessment Results

A total of 39 studies were initially included in our systematic review,[2-40] however only 8 contained sufficient data to inform the recommendation statements and thus underwent formal quality assessment.[6, 15, 27, 29, 30, 35, 37, 40] The remaining 31 studies not incorporated as part of the evidence base are reported in the reference list, but not further described or discussed.[2-5, 7-14, 16-26, 28, 31-34, 36, 38, 39]

In the following sections, the quantity of the evidence as determined by the number of studies that met our inclusion criteria and were retained, the evidence type as determined by study design, the quality of that evidence as determined by the risk of bias assessment, and its consistency are all reported, both as individual studies and in totality, statement by statement.

Statements 1-3 and 5-11 had insufficient published data to inform the recommendations.

Statement 4 is supported by a total of 8 studies that assessed reproducibility, concordance and/or observer agreement.[6, 15, 27, 29, 30, 35, 37, 40] Refer to Supplemental Table 2 for the quality assessment (QA) results for studies informing Statement 4.

Overall, the body of evidence included in this clinical practice guideline represents a summary of the available evidence with a risk of bias of individual studies ranging from low to high. Three of the included studies were determined to have a low risk of bias,[6, 15, 40] one a moderate risk of bias,[37] and four a high risk of bias.[29, 30, 35, 40]

Assessing the Strength of Recommendations

The central question that the panel addressed in developing the guideline was “what procedural principles must be followed in order to assure that HER2 IHC quantitative analysis is accurate and reproducible?”

Development of recommendations required that the panel review the identified evidence and make a series of key judgments:

1. What are the significant findings related to each KQ or outcome? Determine any regulatory requirements and/or evidence that support a specific action.
2. What is the overall strength of evidence supporting each KQ or outcome? Strength of evidence is graded as Convincing, Adequate or Inadequate, based on four published criteria (Supplemental Table 1). Strength of evidence is a key element in determining the strength of a recommendation.
3. What is the strength of each recommendation? There are many methods for determining the strength of a recommendation based on the strength of evidence and the magnitude of net benefit or harm. Supplemental Table 3 describes the ratings for the strength of recommendation, based on the strength of evidence and the likelihood that further studies will change the conclusions. Recommendations not supported by evidence (i.e., evidence was missing or insufficient to permit a conclusion to be reached) were made based on consensus expert opinion. Another potential consideration is the likelihood that additional studies will be conducted that fill gaps in knowledge.

4. What is the net balance of benefits and harms?

Discussion of Benefits and Risks/Harms of Implementing the Recommendations

1. Expert Consensus Opinion. – Laboratories that choose to implement quantitative image analysis (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 (ASCO) and the College of American Pathologists (CAP) guideline "Recommendations for Human Epidermal Growth Factor 2 Testing in Breast Cancer."

By implementing this guideline statement, laboratories using QIA would be compliant with accreditation requirements for validating their imaging system. Successful validation, however, requires proper planning, documentation, and staffing. The laboratory director along with other laboratory personnel will need the expertise required to validate the imaging system for HER2 IHC and may need support from vendors or others outside of their institution who have more experience with the imaging system/algorithm.

The ASCO/CAP HER2 guideline is widely accepted by pathologists. Following the reporting schema of would provide consistency in reporting. This may help with the acceptability of QIA by treating clinicians.

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 immunohistochemistry (IHC).

Results of the QIA must be validated according to laboratory accreditation requirements. Validating QIA results by comparing them to an alternative, validated method will demonstrate the validity of the system. While this may be done in various ways, using FISH or consensus images are offered as acceptable methods.

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

Implementing this recommendation will ensure that laboratories are meeting accreditation requirements and should not result in additional burden to the laboratory. Demonstrating reproducibility will likely increase the laboratory's confidence in the use of QIA for diagnostic purposes.

4. Recommendation. – Laboratories should ensure that the results produced by a quantitative image analysis system are reproducible between operators when they select regions of interest for analysis and/or perform annotation.

Implementing this recommendation will ensure that laboratories are meeting accreditation requirements and should not result in additional burden to the laboratory. Demonstrating reproducibility will likely increase the laboratory's confidence in the use of QIA for diagnostic purposes.

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

Implementing this recommendation will ensure that laboratories are meeting accreditation requirements. Monitoring and documenting performance of laboratory tests/systems is common laboratory practice and this recommendation should not result in additional burden to the laboratory.

6. Recommendation. – Laboratories should have procedures in place to address changes to the quantitative image analysis system that could impact clinical results.

Implementing this recommendation will ensure that laboratories are meeting accreditation requirements. Change procedures are common within the laboratory and this recommendation should not result in additional burden to the laboratory.

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

Though implementing this guideline statement adds more information to the pathology report, this information is required by laboratory accreditation programs, specifically the vendor and name of the imaging system. Any person that sees the report (from the treating physician to administrative staff) will be aware that QIA used. This information could also be used as evidence for QIA billing.

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

Implementing this recommendation ensures laboratories are meeting accreditation requirement. This, in turn, should increase the laboratory's confidence in the personnel and overall use of QIA for diagnostic purposes.

While training in QIA is necessary, laboratories have the burden of defining what this training entails specifically. Vendors may offer a formal training program (i.e., training that involves a certificate of completion or competency), but laboratories should also determine if any additional training/education is necessary beyond the vendor. Developing this training can be time consuming initially, but the panel believes that the benefits of doing so outweigh any harms/risks.

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

The main risk of implementing this guideline statement is that laboratories will have to store results and algorithm metadata in compliance with local requirements and applicable regulations, which may be costly (although necessary for quality control and quality assurance purposes).

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

Implementing this recommendation ensures laboratories are meeting accreditation requirement. This, in turn, should increase the laboratory's confidence in the use of QIA for clinical practice. This recommendation intentionally does not define what "appropriate" means; laboratories should determine the level of expertise the pathologist overseeing this process should have. Laboratories should consider their pathologists' experience/knowledge when determining who will oversee the HER2 QIA process. This pathologist should have advanced expertise in QIA and is able to trouble-shoot in comparison to pathologists who use QIA to sign out HER2 IHC.

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

While the pathologists finalizing the case may not be the same person as the pathologist in charge of overseeing the process, it is important that the former is knowledgeable in the use of the system. The benefit in this knowledge is that he/she should be able to understand if anything he/she is seeing is askew, the clinical validity, and ability to resolve discrepancy.

Dissemination Plans

The CAP plans to host a Quantitative Image Analysis of HER2 IHC for Breast Cancer resource page which will include a link to the manuscript and supplement; a summary of the recommendations, a teaching PowerPoint slide deck (Microsoft Corporation, Redmond, WA), a frequently asked question (FAQ) document, an infographic, and a glossary (also provided at the end of this document). The guideline will be promoted and presented at various society meetings.

Glossary

Accuracy - The degree of correctness or true values of a given laboratory result comparing to a gold standard. Accuracy also implies freedom from error.[41]

Algorithm - A sequential set of instructions used in calculations or problem solving. A diagnostic algorithm or a therapeutic algorithm consists of a stepwise series of instructions with branching pathways to be followed to assist a physician in coming to a diagnosis or deciding on a management strategy, respectively.[42] Image analysis algorithms are used to assist image-based assessment of digital pathology slides.

Batch analysis - An analysis in which all of the samples collected for a specific, nonemergent assay undergo the same testing process at the same time or sequentially.[43]

Function Check – Confirmation that an instrument or item of equipment operates according to manufacturer's specifications before routine use, at prescribed intervals, or after minor adjustment. Depending on the type of system, function checks may include calibration.[44]

Instrument calibration - A process of comparison in which an instrument is used to measure or is measured by a calibration standard, and the result is compared to two things: the known value and uncertainty of the standard and the performance specifications required by the customer. It quantifies the relationship between the readings of a measurement standard under controlled and specified conditions.[45]

Inter-run reproducibility - Also known as within-run precision. Replicates of the same sample across different run demonstrate high correlation.[46]

Intra-run reproducibility - Replicates of samples demonstrate high correlation in the same run. This is to monitor sample to sample variation.[46]

Laboratory developed test (LDT) - A type of in vitro diagnostic test that is designed, manufactured and used within a single laboratory according to the laboratory's own procedures.[47]

Metadata - Data/information that provides information about other data which include descriptive, structural and administrative metadata.[48]

Precision – The closeness of agreement between independent results of measurements obtained under stipulated conditions.[41]

Quantitative image analysis (QIA) – 1) A process whereby quantitative and meaningful information is acquired from the digital images of a specimen, and
2) The computer-assisted detection or quantification of specific features in an image following enhancement and processing of that image, including analysis of immunohistochemistry samples, DNA analysis, morphometric analysis, and *in situ* hybridization.[44]

Quantitative image analysis (QIA) system – A system that integrates automated microscopy, high-quality image acquisition, and powerful analytical algorithms to detect, count, and quantify

areas of interest and includes the hard-wares (computer, scanner, monitor, informatics network, etc.) as well as soft-wares (algorithms).[49]

Region of interest (ROI) - Also called a potential target or image-subregion; a portion of an image that is of interest and upon which an image analysis will be performed on.[50]

Reproducibility - The ability of a test or study to be duplicated either by the same researcher or by someone else working independently.[51]

Revalidation – A procedure used to assess a previously validated test's accuracy and reliability in detecting the marker of interest when there has been a change in test conditions (e.g. methods, equipment, or specimen or fixative types). The degree of revalidation required depends on the nature of the change.

Validation - A documented program that provides a high degree of assurance that a specific process, system or test method will consistently produce a result that accomplishes its intended purpose, meeting predetermined acceptance criteria.[52]

Verification - The process of evaluating a test or a product of a development phase to determine whether it meet the specified requirements. In the post-development phase, verification procedures to endure the product continuously meeting the initial design specifications.[52]

Supplemental Table 1: Grades for Strength of Evidence

Designation	Description	Quality of Evidence
Convincing	High confidence that available evidence reflects true effect. Further research is very unlikely to change the confidence in the estimate of effect.	High/Intermediate quality evidence
Adequate	Moderate confidence that available evidence reflects true effect. Further research is likely to have an important impact on the confidence in estimate of effect and may change the estimate.	Intermediate/Low quality of evidence
Inadequate	Little confidence that available evidence reflects true effect. Further research is very likely to have an important impact on the	Low/Insufficient evidence and expert panel uses formal consensus process to reach recommendation

	confidence in the estimate of effect and is likely to change the estimate.	
Insufficient	Evidence is insufficient to discern net effect. Any estimate of effect is very uncertain.	Insufficient evidence and expert panel uses formal consensus process to reach recommendation

Adapted from J Clin Epidemiol, 64(4), Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence, p. 401-406, copyright 2011, with permission from Elsevier.[53]

Supplemental Table 2: Quality Assessment of Studies Informing Statement 4

Author, year	Patient selection – consecutive or Random	Blinding	Did all pts receive the same reference standard	Appropriate statistical analysis	Funding source	Overall Potential Risk of Bias*
Bolton et al,[6] 2010	NR	√	√	√	Funding from non-profit organizations	Low
Gavrielides et al,[15] 2011	NR	√	√	√	NR, but report no COIs	Low
Gustavson et al,[40] 2009	√	NR	√	√	All authors are employees of HistoRx, Inc (New Haven, CT), the sole licensee of the AQUA technology	High
Minot et al,[27] 2012	√	√	√	√	NR, but reported no COIs	Low
Nassar et al,[29] 2011	NR	√	√	√	Authors paid by industry	High
Prasad et al,[30] 2011	NR	NR	√	√	NR	High
Slodkowska et al,[35] 2010	NR	NR	√	√	Industry	High
Turashvili et al,[37] 2009	NR	NR	√	√	NR, but reported no COIs	Moderate

COI – conflict of interest; NR – not reported;

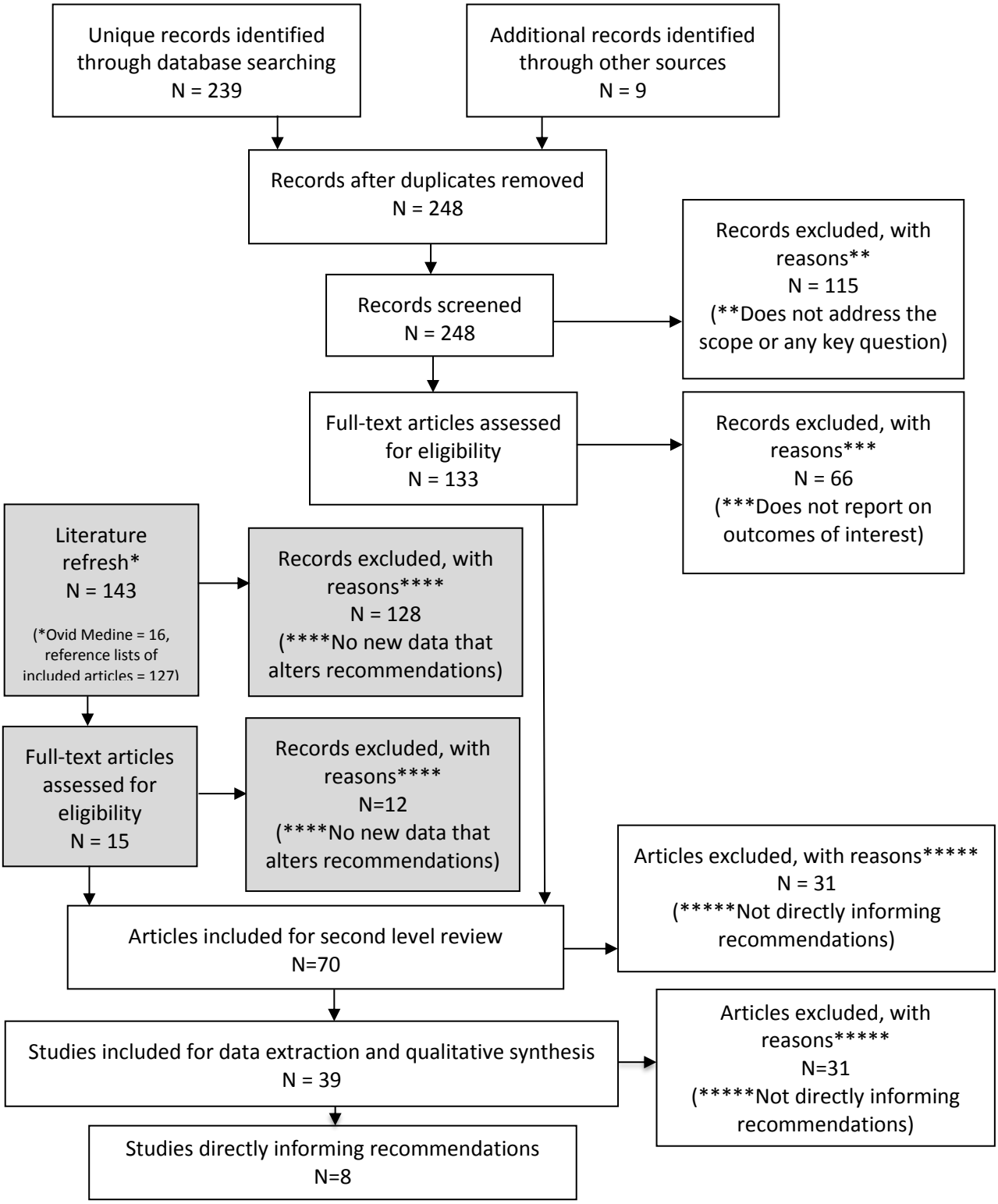
Supplemental Table 3: Grades for Strength of Recommendations

Designation	Recommendation	Rationale
Strong Recommendation	Recommend For or Against a particular practice (Can include “must” or “should”)	Supported by convincing (high) or adequate (intermediate) quality of

		evidence and clear benefit that outweighs any harms
Recommendation	Recommend For or Against a particular practice (Can include “should” or “may”)	Some limitations in quality of evidence (adequate [intermediate]), balance of benefits and harms, values, or costs but panel concludes that there is sufficient evidence to inform a recommendation
Expert Consensus Opinion	Recommend For or Against a particular practice (Can include “should” or “may”)	Serious limitations in quality of evidence (inadequate [low] or insufficient), balance of benefits and harms, values or costs, but panel consensus is that a statement is necessary
No Recommendation	No Recommendation For or Against a particular practice	Insufficient evidence to provide a recommendation, balance of benefits and harms, values or costs

Derived from Andrews et al.[54]

Supplemental Figure 1: Literature Review Flow Diagram *



*Moher D, Liberati A, Tetzlaff J, Altman DG; The PRISMA Group. Preferred reporting items for systematic reviews and meta-analysis: the PRISMA statement. PLoS Med. 2009;6(6):e1000097. doi:10.1371/journal.pmed1000097[55]

Supplemental Figure 2:

Ovid Search String:

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1946 to Present>, Ovid MEDLINE(R) Daily Update <March 17, 2016> Search Strategy:

---

1 image processing, computer-assisted/ (100106)
2 image interpretation, computer-assisted/ (35776)
3 diagnosis, computer-assisted/ (19794)
4 ((quantitat$ or quantif$ or count$ or score or scoring or assess$ or evaluat$ or analyze or analys?s or algorithm$) and (image$ or digital or optical or "whole slide" or WSI)).tw. (326196)
5 or/1-4 (427212)
6 Genes, erbB-2/ (2786)
7 Receptor, ErbB-2/ (18170)
8 (HER?2$ or ERBB?2 or HER2 or HER-2 or ERBB2 or ERBB-2).tw. (27694)
9 "human epidermal growth factor receptor 2".tw. (3664)
10 ERBB2 protein, human.nm. (5597)
11 or/6-10 (31672)
12 exp immunohistochemistry/ (540858)
13 (immunocytochemistry or immunocytochemical or IHC or immunoreactivity or immunohistochemical or immunohistochemistry).tw. (358296)
14 (immunohistochemically or immunocytochemically).tw. (32257)
15 immunofluorescence.tw. (91369)
16 or/12-15 (775226)
17 5 and 11 and 16 (332)
18 remove duplicates from 17 (330)
19 limit 18 to english language (322)
20 limit 19 to yr="2006 -Current" (205)
21 animals/ not humans/ (4172976)
22 20 not 21 (197)
23 limit 22 to (comment or editorial or letter) (0)
24 22 not 23 (197)

PubMed Search String:

(((((((quantitative[tiab] OR quantitate[tiab] OR quantify[tiab] OR count[tiab] OR counting[tiab] OR counts[tiab] OR score[tiab] OR scores[tiab] OR scoring[tiab] OR assess[tiab] OR assessment[tiab] OR evaluate[tiab] OR evaluation[tiab] OR evaluates[tiab] OR analyze[tiab] OR analysis[tiab] OR analyses[tiab] OR algorithm[tiab] OR "algorithms"[MeSH Terms]) AND (image[tiab] OR images[tiab] OR digital[tiab] OR optical[tiab] OR "whole slide"[tiab] OR WSI[tiab]) OR ("image processing, computer-assisted"[MeSH Terms:noexp] OR "image interpretation, computer-assisted"[MeSH Terms:noexp] OR "diagnosis, computer-assisted"[MeSH Terms:noexp])))) AND (("Genes, erbB-2"[MeSH Terms] OR "Receptor, ErbB-2"[MeSH Terms] OR "ERBB2 protein, human"[Supplementary Concept]) OR "human epidermal growth factor receptor 2"[tiab] OR HER-2[tiab] OR HER2[tiab] OR ERBB2[tiab] OR ERBB-2[tiab])) AND (immunocytochemistry[tiab] OR immunocytochemical[tiab] OR IHC[tiab] OR immunoreactivity[tiab] OR immunohistochemical[tiab] OR immunohistochemistry[tiab] OR immunohistochemically[tiab] OR immunocytochemically[tiab] OR immunofluorescence[tiab] OR "immunohistochemistry"[MeSH Terms])) AND ("2006/01/01"[PDat] : "2016/12/31"[PDat]))) NOT ((animals[mh]) NOT humans[mh])

Scopus Search String:

(TITLE-ABS-KEY((immunocytochemistry OR immunocytochemical OR IHC OR immunoreactivity OR immunohistochemical OR immunohistochemistry OR immunohistochemically OR

immunocytochemically OR immunofluorescence))) AND (TITLE-ABS-KEY((erbb-2 OR erbb2 OR her-2 OR her2 OR her-2/neu OR "human epidermal growth factor receptor 2" OR her2/neu))) AND (((TITLE-ABS-KEY(quantitative OR quantitate OR quantify OR count OR counting OR counts OR score OR scores OR scoring OR assess OR assessment OR evaluate OR evaluation OR evaluates OR analyze OR analysis OR analyses OR algorithm OR algorithms)) and (TITLE-ABS-KEY(image or images or digital or optical or "whole slide" or WSI))) or ((TITLE-ABS-KEY("computer assisted") AND TITLE-ABS-KEY("image interpretation")))) AND (LIMIT-TO(PUBYEAR,2016) OR LIMIT-TO(PUBYEAR,2015) OR LIMIT-TO(PUBYEAR,2014) OR LIMIT-TO(PUBYEAR,2013) OR LIMIT-TO(PUBYEAR,2012) OR LIMIT-TO(PUBYEAR,2011) OR LIMIT-TO(PUBYEAR,2010) OR LIMIT-TO(PUBYEAR,2009) OR LIMIT-TO(PUBYEAR,2008) OR LIMIT-TO(PUBYEAR,2007) OR LIMIT-TO(PUBYEAR,2006)) AND (LIMIT-TO(LANGUAGE,"English")) AND (EXCLUDE(DOCTYPE,"le")) AND NOT DBCOLL(medl)

REFERENCES

1. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009
2. Ali HR, Irwin M, Morris L, et al. Astronomical algorithms for automated analysis of tissue protein expression in breast cancer. *Br J Cancer*. 2013;108(3):602-612. doi:10.1038/bjc.2012.558
3. Al-Kofahi Y, Lassoued W, Grama K, et al. Cell-based quantification of molecular biomarkers in histopathology specimens. *Histopathology*. 2011;59(1):40-54. doi:10.1111/j.1365-2559.2011.03878.x
4. Bartlett JM, Christiansen J, Gustavson M, et al. Validation of the IHC4 breast cancer prognostic algorithm using multiple approaches on the multinational TEAM clinical trial. *Arch Pathol Lab Med*. 2016;140(1):66-74. doi:10.5858/arpa.2014-0599-OA
5. 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. doi:10.1136/jcp.2011.089847
6. 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. doi:10.1158/1055-9965.EPI-09-1023
7. Brugmann A, Eld M, Lelkaitis G, et al. Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains. *Breast Cancer Res Treat*. 2012;132(1):41-49. doi:10.1007/s10549-011-1514-2
8. Cantaloni C, Tonini RE, Eccher C, et al. Diagnostic value of automated Her2 evaluation in breast cancer: a study on 272 equivocal (score 2+) Her2 immunoreactive cases using an FDA approved system. *Appl Immunohistochem Mol Morphol*. 2011;19(4):306-312. doi:10.1097/PAI.0b013e318205b03a
9. Cass JD, Varma S, Day AG, et al. Automated quantitative analysis of p53, cyclin D1, Ki67 and pERK expression in breast carcinoma does not differ from expert pathologist scoring and correlates with clinico-pathological characteristics. *Cancers (Basel)*. 2012;4(3):725-742. doi:10.3390/cancers4030725
10. Chang CY, Chang CW, Huang YC, Ko CC. Automatic breast cancer assessment in her-2/neu immunohistochemistry. *Biomed Eng (Singapore)*. 2014;26(4). https://www.worldscientific.com/doi/abs/10.4015/S101623721440002X. Published March 19, 2014. Accessed July 20, 2018.
11. Ciampa A, Xu B, Ayata G, et al. HER-2 status in breast cancer: correlation of gene amplification by FISH with immunohistochemistry expression using advanced cellular imaging system. *Appl Immunohistochem Mol Morphol*. 2006;14(2):132-137. doi:10.1097/01.pai.0000150516.75567.13
12. 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. doi:10.1097/PAS.0000000000000375

13. Dobson L, Conway C, Hanley A, et al. Image analysis as an adjunct to manual HER-2 immunohistochemical review: a diagnostic tool to standardize interpretation. *Histopathology*. 2010;57(1):27-38. doi:10.1111/j.1365-2559.2010.03577.x

14. Gavrielides MA, Conway C, O'Flaherty N, Gallas BD, Hewitt SM. Observer performance in the use of digital and optical microscopy for the interpretation of tissue-based biomarkers. *Anal Cell Pathol (Amst)*. 2014;2014:157308. doi:10.1155/2014/157308

15. 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. doi:10.1043/1543-2165-135.2.233

16. Gavrielides MA, Masmoudi H, Petrick N, Myers KJ, Hewitt SM. Automated evaluation of HER-2/neu immunohistochemical expression in breast cancer using digital microscopy. Paper presented at: 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; May 14-17, 2008; Paris, France. https://ieeexplore.ieee.org/document/4541119/. Accessed July 19, 2018.

17. Hall BH, Ianosi-Irimie M, Javidian P, Chen W, Ganesan S, Foran DJ. Computer-assisted assessment of the human epidermal growth factor receptor 2 immunohistochemical assay in imaged histologic sections using a membrane isolation algorithm and quantitative analysis of positive controls. *BMC Med Imaging*. 2008;8:11. doi:10.1186/1471-2342-8-11

18. Joshi AS, Sharangpani GM, Porter K, et al. Semi-automated imaging system to quantitate Her-2/neu membrane receptor immunoreactivity in human breast cancer. *Cytometry A*. 2007;71(5):273-285. doi:10.1002/cyto.a.20374

19. Keay T, Conway CM, O'Flaherty N, Hewitt SM, Shea K, Gavrielides MA. Reproducibility in the automated quantitative assessment of HER2/neu for breast cancer. *J Pathol Inform*. 2013;4:19. doi:10.4103/2153-3539.115879

20. Keller B, Chen W, Gavrielides MA. Quantitative assessment and classification of tissue-based biomarker expression with color content analysis. *Arch Pathol Lab Med*. 2012;136(5):539-550. doi:10.5858/arpa.2011-0195-OA

21. Laurinaviciene A, Dasevicius D, Ostapenko V, Jarmalaite S, Lazutka J, Laurinavicius A. Membrane connectivity estimated by digital image analysis of HER2 immunohistochemistry is concordant with visual scoring and fluorescence in situ hybridization results: algorithm evaluation on breast cancer tissue microarrays. *Diagn Pathol*. 2011;6:87. doi:10.1186/1746-1596-6-87

22. 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. doi:10.1007/s13277-015-4327-9

23. Lloyd MC, Allam-Nandyala P, Purohit CN, Burke N, Coppola D, Bui MM. Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: How reliable is it? *J Pathol Inform*. 2010;1:29. doi:10.4103/2153-3539.74186

24. Masmoudi H, Hewitt SM, Petrick N, Myers KJ, Gavrielides MA. Automated quantitative assessment of HER-2/neu immunohistochemical expression in breast cancer. *IEEE Trans Med Imaging*. 2009;28(6):916-925. doi:10.1109/TMI.2009.2012901

25. Micsik T, Kiszler G, Szabo D, et al. Computer aided semi-automated evaluation of HER2 immunodetection--a robust solution for supporting the accuracy of anti HER2 therapy. *Pathol Oncol Res*. 2015;21(4):1005-1011. doi:10.1007/s12253-015-9927-6

26. Minot DM, Kipp BR, Root RM, et al. Automated cellular imaging system III for assessing HER2 status in breast cancer specimens: development of a standardized scoring method that correlates with FISH. *Am J Clin Pathol*. 2009;132(1):133-138. doi:10.1309/AJCPJV0SKAF2PCMY

27. 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. doi:10.1309/AJCP9MKNLHQNK2ZX

28. Mohammed ZM, Going JJ, McMillan DC, et al. Comparison of visual and automated assessment of HER2 status and their impact on outcome in primary operable invasive ductal breast cancer. *Histopathology*. 2012;61(4):675-684. doi:10.1111/j.1365-2559.2012.04280.x

29. 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. doi:10.1043/2010-0418-OAR1.1

30. 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. doi:10.5306/wjco.v2.i4.187

31. Prasad K, Zimmermann B, Prabhu G, Pai M. Datamining approach for automation of diagnosis of breast cancer in immunohistochemically stained tissue microarray images. *Open Med Inform J*. 2010;4:86-93. doi:10.2174/1874431101004010086

32. 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. doi:10.5858/arpa.2014-0127-OA

33. Skaland I, Ovestad I, Janssen EA, et al. Digital image analysis improves the quality of subjective HER-2 expression scoring in breast cancer. *Appl Immunohistochem Mol Morphol*. 2008;16(2):185-190. doi:10.1097/PAI.0b013e318059c20c

34. Skaland I, Ovestad I, Janssen EA, et al. Comparing subjective and digital image analysis HER2/neu expression scores with conventional and modified FISH scores in breast cancer. *J Clin Pathol*. 2008;61(1):68-71. doi:10.1136/jcp.2007.046763

35. 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. doi:10.2478/v10042-010-0015-1

36. Tawfik OW, Kimler BF, Davis M, et al. Comparison of immunohistochemistry by automated cellular imaging system (ACIS) versus fluorescence in-situ hybridization in the evaluation of HER-2/neu expression in primary breast carcinoma. *Histopathology*. 2006;48(3):258-267. doi:10.1111/j.1365-2559.2005.02322.x

37. 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. doi:10.1186/1471-2407-9-165

38. 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. doi:10.1371/journal.pone.0123201

39. von Bueren E, Patel N, McGann-Tiedt K, Sandbank B, Muralitharan SS. Analysis of HER2 peptide - based analyte control slides. *J Histotechnol*. 2010;33(4):153-156.

40. 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. doi:10.1043/1543-2165-133.9.1413

41. Wagar EA. Laboratory laws and regulations. In: Wagar EA, Horowitz RE, Siegal GP, eds. *Laboratory Administration for Pathologists*. Northfield, IL: College of American Pathologists Press; 2011:215-234.

42. Martin E. Oxford Concise Medical Dictionary, 9th ed. Oxford, United Kingdom: Oxford University Press; 2015. http://www.oxfordreference.com/abstract/10.1093/acref/9780199687817.001.0001/acref-9780199687817-e-282?rskey=lZze6d&result=321. Accessed July 15, 2018.

43. Venes D, ed. *Tabers Cyclopedic Medical Dictionary*. 23rd ed. Philadelphia, PA: F. A. Davis Company; 2017. https://www.tabers.com/tabersonline/view/Tabers-Dictionary/743387/0/batch_analysis?q=analysis+batch. Accessed July 15, 2018.

44. College of American Pathologists. Accreditation checklists. College of American Pathologists website. https://www.cap.org/laboratory-improvement/accreditation/accreditation-checklists. Accessed July 16, 2018.

45. Payne GC. Calibration: What Is It? *Quality Progress*. asq.org/quality-progress/2005/05/measure-for-measure/calibration-what-is-it.html. Published May 2005. Accessed July 16, 2018.

46. Chesher D. Evaluating assay precision. *Clin Biochem Rev*. 2008;29(Suppl 1):S23-26.

47. U.S. Food & Drug Administration. Laboratory Developed Tests. U.S. Food & Drug Administration website. Updated March 26, 2018. Accessed July 16, 2018.

48. Riley J. *Understanding Metadata: What is metadata, and what is it for? A Primer*. Baltimore, MD: NISO Press; 2017.

49. van der Hall PO, Kraan MD, Tak PO. Quantitative image analysis of synovial tissue. *Methods Mol Med*. 2007;135:121-143.

50. Si J, Lin H, Abousleman G. Electronic imaging and signal processing: detecting regions of interest in images. *SPIE Newsroom*. http://spie.org/newsroom/0414-detecting-regions-of-interest-in-images?SSO=1. Published November 2006. Accessed July 18, 2018.

51. Samsa G, Samsa L. A guide to reproducibility in preclinical research [published online July 10, 2018]. *Acad Med*. doi: 10.1097/ACM.0000000000002351

52. U.S. Department of Health and Human Services, Food and Drug Administration. Quality System Regulation, General Provisions, Definitions. 21 CFR §820.3 2012.

53. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-406. doi:10.1016/j.jclinepi.2010.07.015

54. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7):719-725. doi:10.1016/j.jclinepi.2012.03.013

55. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097