

Original Article

Whole slide imaging for human epidermal growth factor receptor 2 immunohistochemistry interpretation: Accuracy, Precision, and reproducibility studies for digital manual and paired glass slide manual interpretation

David C. Wilbur, Elena F. Brachtel, John R. Gilbertson, Nicholas C. Jones¹, John G. Vallone², Savitra Krishnamurthy³

Department of Pathology, Harvard Medical School, Massachusetts General Hospital, ¹Department of Pathology, Massachusetts General Hospital, Boston, MA, ²Department of Pathology, University of Southern California, Los Angeles, California, ³Department of Pathology, MD Anderson Cancer Center, Texas, USA

E-mail: *Dr. David C. Wilbur - dwilbur@partners.org

*Corresponding author

Received: 29 November 14

Accepted: 03 March 2015

Published: 28 May 2015

This article may be cited as:

Wilbur DC, Brachtel EF, Gilbertson JR, Jones NC, Vallone JG, Krishnamurthy S. Whole slide imaging for human epidermal growth factor receptor 2 immunohistochemistry interpretation: Accuracy, Precision, and reproducibility studies for digital manual and paired glass slide manual interpretation. *J Pathol Inform* 2015;6:22.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2015/6/1/22/157788>

Copyright: © 2015 Wilbur DC. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Background: The use of digital whole slide imaging for human epidermal growth factor receptor 2 (HER2) immunohistochemistry (IHC) could create improvements in workflow and performance, allowing for central archiving of specimens, distributed and remote interpretation, and the potential for additional computerized automation. **Procedures:** The accuracy, precision, and reproducibility of manual digital interpretation for HER2 IHC were determined by comparison to manual glass slide interpretation. Inter- and intra-pathologist reproducibility and precision between the glass slide and digital interpretations of HER2 IHC were determined in 5 studies using DAKO HercepTest-stained breast cancer slides with the Philips Digital Pathology System. In 2 inter-method studies, 3 pathologists interpreted glass and digital slides in sequence or in random order with a minimum of 7 days as a washout period. These studies also measured inter-observer reproducibility and precision. Another two studies measured intra-pathologist reproducibility on cases read 10 times by glass and digital methods. One additional study evaluated the effects of adding IHC control slides with each run, using 1 pathologist interpreting glass and digital slides randomized from the sets above along with appropriate controls for each slide in the set. **Results:** The overall results show that there is no statistical difference between the variance of performance when comparing glass and digital HER2 interpretations; and there were no effects noted when control tissues were evaluated in conjunction with the test slides. **Conclusions:** The results show that there is an equivalence of result when interpreting HER2 IHC slides in breast cancer by either glass slides or digital images. Digital interpretation can therefore be safely and effectively used for this purpose.

Key words: Digital pathology, human epidermal growth factor receptor 2 immunohistochemistry, whole slide imaging

Access this article online

Website:
www.jpathinformatics.org

DOI: 10.4103/2153-3539.157788

Quick Response Code:



INTRODUCTION

The use of digital imaging technology has been increasing in a wide variety of applications in recent years. Use for permanently archiving rare specimens or consultations, educational programs, and telepathology have all been reported.^[1-3] Whole slide digital images have the potential to replace glass slides as the predominant viewing modality by the pathologist for diagnostic purposes. This paradigm shift has the potential to impart very important improvements to workflow, cost, accuracy, and overall healthcare outcomes. Whole slide images represent digital reproductions of entire glass slide histopathology preparations, and if captured at high enough resolution by a system that maintains and displays the images with fidelity comparable to light microscopy should perform in an equivalent fashion to routine glass slides. A number of applications using whole slide images in a variety of specimen types have shown comparability of interpretation between whole slide images and glass slides. These include the primary diagnosis, frozen section interpretation, second opinion consultations, and the interpretation of special studies, like immunohistochemistry (IHC).^[1-3]

Other studies have pointed out important caveats to digital interpretation, which include the need to vet systems for intended use situations. These include the slide scanner, image management systems, as well as the interfaces that pathologists use to review the captured and stored images. The US Food and Drug Administration (FDA) has suggested that such systems require regulation (particularly for primary interpretations), although they have not yet issued guidance in this regard. The College of American Pathologists convened a consensus process to evaluate the procedures that should be followed in the validation of this technology if used for primary interpretations where no glass slide-based routine microscopy is used. The principles of validation include adequate numbers of comparison challenges between the 2 methods with no statistical difference noted in error rates.^[4]

There have been a number of reports comparing the accuracy of the interpretation of immunohistochemical preparations using routine glass slide and digital imaging.^[5-7] Some have indicated that results are comparable between the 2 methods and others have reported differences.

Human epidermal growth factor receptor 2 (HER2) IHC is important in assessing who will be responsive to targeted therapy using drugs directed at the epidermal growth factor receptor. As such, an evaluation of HER2 is now a routine part of all breast cancer workups and the reporting of results has become highly systematized due to the need for accurate and reproducible results. Unlike

many IHC assays, HER2 requires precise scoring in order to be effective and hence reproducibility studies are necessary to insure safe and effective operation of a digital scoring method. It is estimated that approximately 20% of current manual glass slide-based HER2 interpretation may be inaccurate.^[8] Early studies using digital imaging for the evaluation of selected fields (not whole slide images) with computer-assisted scoring for HER2 assessment showed that use of this technique could improve the concordance (and hence the overall accuracy) between the immunohistochemical HER2 result and the subsequent fluorescence *in situ* hybridization result.^[6,7,9] More recent studies have confirmed these results, but have noted that there are caveats related to scanning equipment that do require validation of each system that is intended for clinical use.^[10] The use of whole slide images to assess HER2 IHC has been reported,^[11] and in another study using 3 levels of result (negative [0-1+], equivocal [2+], and positive [3+]), showed substantially equivalent percent agreement (PA) for inter- and intra-observer reproducibility with these figures ranging from 61% to 92%.^[12] To date, several commercial vendors have sought and received FDA clearance for the manual digital interpretation of HER2 via whole slide image review; however the FDA considers prior approvals on different instrumentation to be insufficient to assume equivalence for all systems, and hence each requires its own studies for validation.^[13,14]

The purpose of the present study was to assess the performance of the Philips Digital Pathology Solution™ (Philips Digital Pathology, Best, The Netherlands) in a protocol designed for FDA submission. As a by-product, the data obtained builds on the results from the previously noted studies by assessing the intermethod PA in a variety of settings obtained from routine glass slide and whole slide image interpretations of HER2 IHC in breast cancer cases.

MATERIALS AND METHODS

The study was conducted with the approval of the appropriate institutional review boards. There were 5 separate studies conducted to assess the safety and effectiveness of HER2 IHC interpretation by digital pathology methods.

General Methods (Apply to all Studies Below)

All IHC slides used in the study were selected from cases of known breast cancer within the Dako (Glostrup, Denmark) tissue bank and were stained with the Dako HercepTest™ as previously described.^[15] Slide sets were constructed to include all score categories with an equal score distribution, as determined by an independent pathologist using optical interpretations, in order to reduce bias toward any particular result type. Training was provided to all participating pathologists with standard

sets presented and discussed at a multi-head microscope and via the digital viewing platform. Demonstrated competency with HER2 scoring and with operation of the digital interface was verified prior to study initiation. The pathologists performing the readings were masked to the slide set construction criteria. All digital pathology specimens were generated using the Philips Ultra Fast Scanner [Figure 1a], stored in the Philips Image Management System, and viewed on the Philips image viewer [Figure 1b]. HER2 slides (both manual glass and manual digital) were scored (0, 1+, 2+, or 3+) according to the HerceptTest kit labeling package insert.^[15] No automated image analysis was performed in this protocol. For the purposes of primary comparison, the HerceptTest scores were grouped into negative (0 and 1+) and positive (2+ and 3+) as this would represent a cutoff which would define a tentative positive result requiring further HER2 testing, or a positive result, respectively (binary correlations). In addition, exact trichotomous correlations (3×3 correlations) by discrete scoring category (combining 0 and 1+, and leaving 2+ and 3+ uncombined) were collected and are reported, but were not used as part of the acceptance criteria for inter-method equivalence. In addition, for slides that were read multiple times during different phases of the study, the washout time period between reviews was captured and analyzed to see if there was any correlation effect based on the length of the washout period. Because there were pools of cases that had and had not been read multiple times during the course of the study a *post-hoc* analysis was performed in order to determine if there was a difference in the concordance rates between these pools. "Acceptance criteria" mentioned below were adopted in association with US FDA guidance as indicative of study success.

Inter-method Studies (2 Studies)

The primary objective of these studies was to test

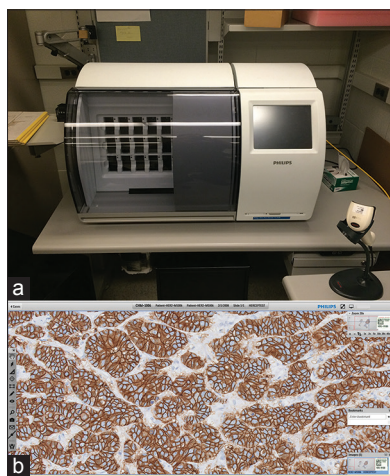


Figure 1: Images of Philips ultrafast scanner (a) and Philips viewer with human epidermal growth factor receptor 2 immunohistochemistry display (b)

the hypothesis that there is no difference in manual interpretation of the cases between the glass slide review and digital whole slide image review. In order to meet this goal, in the first study, the binary overall, positive and negative inter-pathologist and inter-method PA needed to be equal or better than the comparable agreements for glass slide review. In the second study, a threshold of >75% for inter-pathologist binary overall PA and inter-method binary overall, positive and negative PA (PPA and NPA) was required over all slides in the sets. In the first study, 3 pathologists interpreted HER2 IHC cases first by glass slide review in a standard microscope and then by a digital process (195 cases). In the second study, 3 pathologists (same 2 pathologists who participated in study 1 and 1 additional pathologist) interpreted HER2 IHC cases by glass slide and digital methods with the order of first and second method reads determined by randomization (200 cases, of which 73 cases were the same cases as used in study 1 (Note: To prevent recall bias these cases had been re-randomized and re-labeled).

In both studies, there was a minimum of a 7-day washout period between any paired reading. The order of review of the paired cases was randomized for each pathologist. Inter-pathologist agreement was determined for each pair of pathologists for glass slide versus glass slide, glass slide versus digital, and digital versus digital. In both studies, the pathologists were masked to the slide set construction or to any prior results.

Intra-pathologist and Inter-pathologist Studies (2 Studies)

Two studies, each with one pathologist, interpreted 2 different HER2 immunohistochemical slide sets containing 8 test slides randomly mixed with 12 non test slides which differed per read and were added to diminish bias. The pathologists in each study were masked to the set construction and to any prior results. Each set of 20 slides was interpreted 10 times (5 times each by the glass slide and digital methodology) with the order of cases and the reading method randomized for each set. Intra-pathologist agreement (reproducibility) was determined for each pathologist. Inter-pathologist agreement was determined by comparisons of the interpretation of each pathologist and the prior reads on the same cases in the initial 2 inter-method studies above ($n = 73$). Some, but not all of the slides used in this study were the same cases as had been used previously.

Inter-method Reproducibility Study Using, Human Epidermal Growth Factor Receptor 2 Control Immunohistochemistry Slides

To analyze any possible effect of having run-specific HER2 IHC control tissues to examine in conjunction with assessment of the test slides, an additional study was performed. One pathologist interpreted a set of 53 glass

and digital slides randomized from the sets above, with the addition of run-specific immunohistochemical control slides. To prevent recall bias, all slides were relabeled and randomized for order of review. The control slides were reviewed (glass slide or digital as appropriate) in conjunction with each test slide. The pathologist was masked to the set construction or any prior results. This protocol therefore more closely mimicked actual clinical practice for HER2 IHC evaluation. The agreement rate between glass and digital reads was determined. The hypothesis to be tested was that there would be no difference between the scoring of slides with and without the review of control slides for both glass slide and digital reviews.

Statistical Methods

For inter-method comparisons, statistical analyses were provided overall and per pathologist for a dichotomous categorization of HercepTest™ scores combining 0 and 1+ (negative results) and combining 2 + and 3+ (positive results). For the inter-pathologist analysis, pairwise PA estimates were determined.

For the acceptance criteria evaluation, overall PA estimates for inter-method comparisons were calculated as the average of PA values respectively across all pathologists. 95% confidence intervals (CIs) for the overall estimates were calculated by the bootstrap method.

Similarly, for the acceptance criteria evaluation, the overall PA estimate for the inter-pathologists (accuracy) comparisons was calculated as the average of PA values from the 3 pairwise PA values. This estimation will be done separately for the manual digital method and for the manual optical methods. A bootstrap 95% CI was constructed in each case.

For the intra-pathologist studies, comparisons were compiled per method. For the control study, the PA and the

corresponding CIs were calculated for manual digital with control slides versus manual digital without control slides, and for manual optical with control slides versus manual optical without control slides. Agreement rates between all paired readings were calculated. The hypothesis in the first study was that the aggregate of concordance between all reads (inter-and intra-pathologist) would be >80% and that this would constitute substantial equivalency. For the second study, the aggregate of concordance between all reads (inter-pathologist and inter-method) would be >75%. In addition, an analysis of “outliers” in each of the arms of the studies was performed. An “outlier” is defined as a score result (0–3+) that is different from the median value of the scores of the pathologists over the total set. The hypothesis (acceptance criteria as determined from historical studies used for FDA clearance^[13,14]) for the inter-pathologist studies was that digital arm outliers would represent <20% of cases in the first and second studies. For intra-pathologist studies, the acceptance criteria for digital outliers would be <10% of cases.

RESULTS

Inter-method Studies

In the first study, in which glass slides were always reviewed first followed by the digital slides, 195 matched cases (glass and digital slides) were reviewed, of which 180 had scores for both methods and all three pathologists. The binary agreement rates for pathologist versus pathologist ranged from 88% to 94% (optical), 86–92% (digital), and 81–92% (inter-method) [Figure 2]. There was no statistical difference in variance between the optical and digital pair agreement rates. Based on the goal of equivalent agreement or better for the digital compared to optical interpretations over the entire set, the criteria for acceptance were met. The overall PA was 87.2%, with

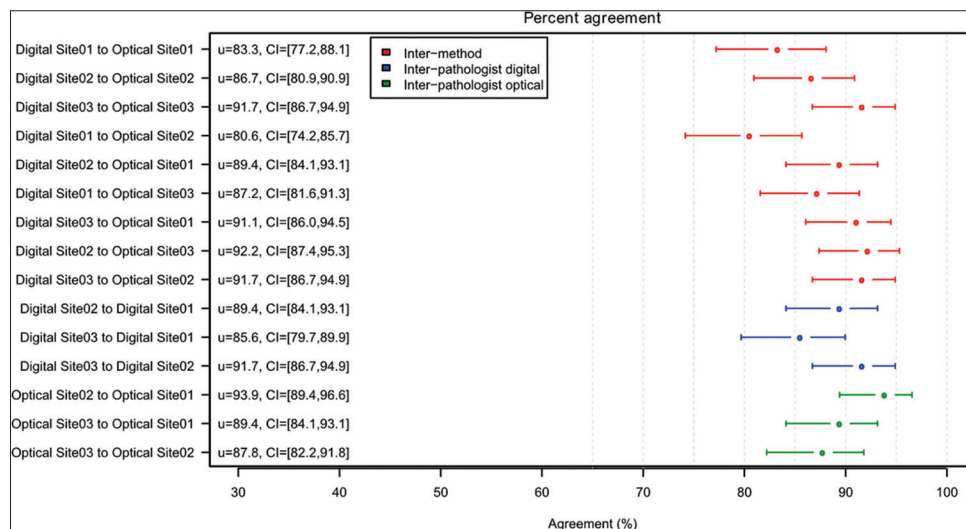


Figure 2: Binary pathologist agreement rates in inter-method study I. The percentage agreements are shown with 95% confidence intervals for each of the comparisons

the PPA being 94.9%, and the NPA being 81.3% [Table 1]. The number of outliers was 14.3% for glass slide review and 15.2% for digital review. The trichotomous (3 × 3) correlations for glass compared to digital ranged from 74.9% to 86.2% (per pathologist, full analysis set per pathologist) and 75.0–86.1% (per pathologist, full analysis set overall). In the second study, in which the order of interpretation was determined by randomization, 200 matched cases (glass and digital slides) were reviewed of which 184 had scores for both methods and all three pathologists. The binary agreement rates for pathologist versus pathologist ranged from 86% to 91% (optical), 78–91% (digital), and 83–92% (inter-method). Based on the goal of > 75% overall agreement in the entire set, the criteria for acceptance was met. The overall PA was 88.8%, with the PPA being 95.7%, and the NPA being 82.8% [Table 1]. Outliers comprised 12.0% for glass slide results and 13.4% for digital results. The trichotomous (3 × 3) correlations for glass slide compared to glass slide ranged from 78.3% to 88.6% (full analysis set overall); for digital compared to digital they ranged from 73.4% to 84.8% (full analysis set overall); for glass compared to digital they ranged from 81.1% to 88.0% (per pathologist, full analysis set per pathologist) and 80.4–88.0% (per pathologist, full analysis set overall).

Intra-pathologist Studies

In the 2 studies using sets of slides (optical and digital) that were interpreted multiple times by an individual pathologist, the paired agreement rates for intra-pathologist reproducibility were as follows: Study 1 (pathologist a) – 89% (optical), 95% (digital); study 2 (pathologist b) – 88% (optical), 93% (digital). The percentages of outliers were 10.8% and 12.5% for the glass slide reviews and 5.1% and 7.5% for the digital reviews in the 2 studies, respectively [Table 2]. For the inter-pathologist reproducibility portions of these studies, where slides interpreted here were compared against the prior interpretations in the 2 studies above, the PA were as follows: 86 and 88% (optical), and 85 and 87% (digital). In both the intra- and inter-pathologist components of these studies, the criteria for acceptance were met.

Control Slide Study

The overall PA in this study was 88.5% for glass slide review and 94.1% for digital review. The PPA for glass slide and digital reviews were 75.0% and 85.7%, and for NPA were 100.0% and 100.0%, respectively [Table 3]. Based on the acceptance criteria, there was no indication of variability between readings when control slides were or were not reviewed as part of the interpretive process.

The Effect of the Washout Period

There was a minimum of a 7-day washout period for all duplicate case reviews, however, the reality of study organization led to longer washout periods in many instances. As shown in Figure 3 the overall

intra-pathologist binary agreement rates did not show any statistically significant differences when measured against washout time periods that ranged from 7 to 51 days.

Concordance Rates for Slide Pools Having Been Read Once or Multiple Times

Table 4 shows the results of the comparison between the binary concordance rates from a common set (which had been read more than once during the study – 73 cases) and a noncommon set (which had been read only once – 114 cases). The 95% CIs for PA, PPA, and NPA are overlapping between the prior and current scorings, indicating that there is no statistical difference between the concordance rates between the 2 pools and hence no identifiable effect of multiple versus single reviews. These results argue that there was no recall bias effect in the slides that had been reviewed multiple times during the course of the studies.

Comment

In this report, the results of multiple studies of inter- and intra-observer concordance show that there is no statistical difference between the interpretation of HER2 IHC

Table 1: The overall PA, the PPA cases and the NPA cases for studies 1 and 2

Study	PA	PPA	NPA
Inter-pathologist (95% CI)	87.2 (84.1-89.8)	94.9 (91.3-97.1)	81.3 (76.6-85.3)
Inter-pathologist (95% CI)	88.8 (85.7-91.7)	95.7 (93.1-98.0)	82.8 (77.9-87.3)

CI: Confidence interval, PA: Percentage agreements, PPA: Positive percentage agreements, NPA: Negative percentage agreements

Table 2: The binary PA from the 2 intra-pathologist studies and the percentage of outliers from each study

Method	Binary PA Outliers	
	Study - 1	Study - 2
Optical manual (%)	89.2 10.8	87.5 12.5
Digital manual (%)	94.9 5.1	92.5 7.5

PA: Percentage agreements

Table 3: The binary PA for cases in which matched control slides were evaluated as a part of the interpretation procedure

Method	Binary PA (%)	95% CI
Optical manual	88.5	77.0-94.6
Digital manual	94.1	84.1-98.0

PA: Percentage agreements, CI: Confidence interval

Table 4: The concordance rate for slides that were read multiple times during the study compared to those that were read a single time. No statistical differences were noted in these 2 slide pools indicating that there was no identifiable recall bias from reading slides multiple times

Slide Set	Intermethod	n	PA (%) (95%CI)	PPA (%) (95% CI)	NPA (%) (95% CI)
Common	Site 3 DigPrev vs OptPrev	68	88 (78-94)	82 (66-91)	94 (81-98)
	Site 1 DigPrev vs OptPrev		88 (78-94)	100 (87-100)	81 (67-90)
	Site 3 DigCur vs Optcur	70	96 (88-99)	100 (88-100)	93 (81-97)
	Site 1 DigCur vs Optcur		90 (81-95)	86 (72-94)	94 (80-98)
Non-Common	Site 3 DigPrev vs OptPrev	112	94 (88-97)	91 (81-96)	96 (88-99)
	Site 1 DigPrev vs OptPrev		86 (78-91)	98 (88-100)	78 (66-86)
	Site 3 DigCur vs Optcur	114	89 (83-94)	95 (83-99)	87 (77-93)
	Site 1 DigCur vs Optcur		92 (86-96)	94 (85-98)	90 (80-95)

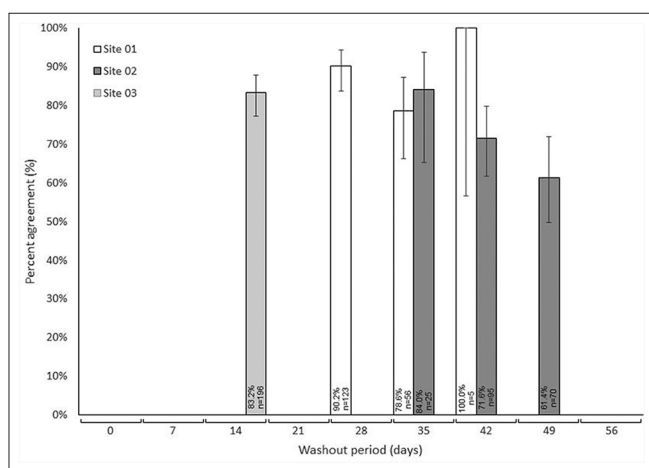


Figure 3: The effect of the length of the washout period between readings is illustrated. No statistical differences between the washout period length and the percent agreements were identified

slides read by the routine glass slides and by digital whole slide imaging (WSI) with the Philips system. In addition, the studies indicate that there was no effect noted when IHC control slides were used or not used as part of the interpretation process, that there was no effect of the length of time of the washout period between readings, and that there was no effect attributable to recall bias on slides that were reviewed multiple times during the study. In these studies, the rate of concordance of readings was determined for glass slide interpretation and the digital interpretation. The rates of agreement between observers reviewing glass slides or digital slides were found to be equivalent for both intra- or inter-method comparisons. Overall, these studies provide substantial evidence that, for the Philips system, the use of digital WSI of HER2 slides for immunohistochemical interpretation is a safe and effective procedure. The results as herein reported led to a US FDA 510K clearance for clinical use of the Philips manual digital HER2 interpretation system (9-19-13).^[16]

The use of WSI for IHC in general and HER2 in particular, is an important potential paradigm shift in the workflow of pathology specimen evaluation. When

considering the broad field of IHC, WSI has already been shown to allow the broadest availability of marker types essentially to every pathologist, regardless of the practice environment. A model in which histology blocks are sent to a centralized facility for IHC staining, followed by WSI scanning and digital transmission back to the source pathologist, allows all practitioners the same access to up-to-date panels of markers, centralized consultative expertise, consistently high quality control, large validation sets, and the opportunity to maintain interpretation expertise and control of professional reimbursement. With the need for very high-quality technical standards inherent in quantitative IHC, extending this concept to HER2 testing is only natural. Centralized facilities could perform the technical portion of the assay, and originating sites can interpret WSI of the scanned specimens, and those interpretations could be reviewed by other pathologists or expert panels for quality assurance. In addition to the quality laboratory standardization achieved, the results of the present study suggest that not only are observers able to accurately score HER2 WSIs, but the data suggest that they may be more consistent in that scoring with WSI than with glass slides. In the 2 intra-observer studies in which a series of slides were read multiple times by both glass slide and digital methods, the digital methods achieved higher (although not statistically different) levels of concordance (pathologist 1–89% vs. 95% and pathologist 2–88% vs. 93%, respectively). Although not studied specifically in this trial, one might speculate that WSI images allow a complete view of the slide at very low magnification, which may allow for a more consistent identification of the best area in which to perform high magnification examination for the final scoring. In addition, the viewing of wider areas at mid or high magnifications allows for a broader view of fields to be scored; this allowing for more consistent and hence improved, overall consistency. Future research might focus on the field selection process and the effects of monitor size and resolution; parameters which may alter the contrast and brightness of the immunohistochemical staining patterns.

These results and accompanying observations may also argue for WSI evaluation of HER2 slides for institutions performing their IHC in-house. Specimens can be easily archived and retrieved for additional or subsequent review. The ability to rapidly rereview a case and compare with a new specimen by pulling from the digital archive is a distinct advantage over glass slides which may be lost, take time to retrieve, or fade with aging. Furthermore, the field upon which the interpretation is made could be marked and stored electronically for subsequent review, a process unlikely to be routinely performed with glass slide analysis. And, digital images can be used for teaching purposes be they local or at a distance.

Because of the need for continuous quality assurance in the HER2 assay it has been suggested that external quality assurance exercises, such as inter-laboratory comparison programs be implemented.^[17-19] The use of digital images for such programs would be substantially more efficient than transferring glass slide assets, and hence the importance of studies such as in the present report which shows that interpretations using digital telepathology technology are equivalent to reading local glass slides.

The addition of image analysis could easily be applied to WSI as has been reported for other systems. The use of image analysis combined with WSI-based field selection has the potential to yet further increase the consistency and precision of interpretation.^[6,7,9,10,20-23] WSI-based HER2 image analysis on the system in this report is currently undergoing preliminary testing, but was not included in this trial.

It is important to obtain these types of system “validation” results prior to the marketing of a specific device for digital interpretation of HER2 to insure safety and effectiveness. HER2 evaluations are an important determinant of the potential utility of targeted therapy in breast cancer. The most recent ASCO guidelines (2014) indicate that HER2 should be evaluated on all invasive breast cancers in order to determine the potential effectiveness of targeted therapy with trastuzumab against amplified HER2 receptor.^[24] There is also considerable importance given to reproducibility of results to insure that this expensive therapy is given to those patients who are most likely to benefit from its use. Although multiple prior studies of immunohistochemical stains in general,^[5,25,26] and more specifically HER2 studies have shown substantial agreement between glass slide and digital interpretations,^[6,7,9,11,12,27] studies of automated image analysis algorithms on HER2 interpretation comparing different scanning devices and different image analysis algorithms have suggested that there may be differences between the instruments and hence that verification of each individual system’s performance should be accomplished prior to clinical use.^[10] Individual

laboratories that use these FDA cleared devices do not have to repeat the studies presented herein, but do have to perform verification procedures to insure that the device/system is performing at its expected level.

Food and Drug Administration 510k clearance has been previously granted for several HER2 digital whole slide image reading systems.^[13,14] The results of 2 of these studies, both with similar protocols to the present studies, showed essentially equivalent results to those noted herein. In the Aperio study (FDA K071671), trichotomous PA between the methods ranged from 61% to 93%, and for the Olympus study (FDA K111914), trichotomous PA ranged from 79% to 93%. In the present study, the trichotomous PA was 81–88%. 95% CIs for the 3 studies were overlapping. The number of cases with score differences of 2 or greater between the readings were comparable between the Olympus and Philips studies (0 and 1 cases, respectively), and these studies were less than with the Aperio study (6 cases) [Table 5].

The CAP guidelines for validation indicate that washout periods are important to prevent recall bias in primary interpretations.^[4] Washout periods and recall bias were not shown to be of significant influence in the present study. It is unlikely that it would be an issue as it might be for diagnostic surgical pathology specimens. Given that scoring is the main output of this HER2 IHC study, memory of individual fields used for the scoring is unlikely and recollection of how any particular field was scored would be difficult to recall. On the other hand, in surgical pathology diagnostic interpretations, the overall specimen is reviewed and a diagnosis is rendered. For the latter types of studies, it is far more likely that there will be a memory of the final diagnosis for the case. Hence, the present result of no recall effect is not unexpected.

In summary, the results of this study are entirely

Table 5: The results of the current studies to similar prior FDA submissions that resulted in clearance. Where the studies were comparable in design there were no statistically significant differences noted

510(k)	Aperio K071671	Olympus K111914	Philips K130020
Lower PA range (%)	61.25	79.00	81.05
Upper PA range (%)	92.50	93.00	88.04
Lower 95% CI range (%)	49.7	70.02	74.89
Upper 95% CI range (%)	97.2	96.57	91.97
MO 0/1+ to MD 2+ (3 largest) (%)	29.79	40.00	31.25
	25.00	36.11	10.66
	13.79	14.81	8.60
Score jumps of 2 (n)	6	0	1

FDA: Food and Drug Administration, PA: Percentage agreements, CI: Confidence interval

consistent with the results of similar studies done in support of FDA clearance and as research protocols. The results show that using the Philips system for the manual scoring of digital whole slide images of HER2 IHC is equivalent to manual glass slide HER2 review and scoring. The study also shows that for this type of IHC assay, the interval of the washout period (beyond 7 days) does not alter the results, and that observers did not appear to be biased by having previously reviewed a given slide. Therefore, using the Philips system for the manual HER2 scoring of digital WSI can be considered safe and effective, and the potential benefits of using digital technology for this assessment as described above could be implemented for this method.

REFERENCES

- Pantanowitz L, Valenstein PN, Evans AJ, Kaplan KJ, Pfeifer JD, Wilbur DC, et al. Review of the current state of whole slide imaging in pathology. *J Pathol Inform* 2011;2:36.
- Weinstein RS, Graham AR, Richter LC, Barker GP, Krupinski EA, Lopez AM, et al. Overview of telepathology, virtual microscopy, and whole slide imaging: Prospects for the future. *Hum Pathol* 2009;40:1057-69.
- Al-Janabi S, Huisman A, Van Diest PJ. Digital pathology: Current status and future perspectives. *Histopathology* 2012;61:1-9.
- Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, et al. Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2013;137:1710-22.
- Rizzardi AE, Johnson AT, Vogel RI, Pambuccian SE, Henriksen J, Skubitz AP, et al. Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol* 2012;7:42.
- Wolff AC, Hammond EH, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 2007;131:43.
- Bloom K, Harrington D. Enhanced accuracy and reliability of HER-2/neu immunohistochemical scoring using digital microscopy. *Am J Clin Pathol* 2004;121:620-30.
- Skaland I, Øvestad I, Janssen EA, Klos J, Kjellevoid KH, Heliessen T, 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;61:68-71.
- 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;135:233-42.
- 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.
- Kondo Y, Iijima T, Noguchi M. Evaluation of immunohistochemical staining using whole-slide imaging for HER2 scoring of breast cancer in comparison with real glass slides. *Pathol Int* 2012;62:592-9.
- Nassar A, Cohen C, Albitar M, Agersborg SS, Zhou W, Lynch KA, et al. Reading immunohistochemical slides on a computer monitor – A multisite performance study using 180 HER2-stained breast carcinomas. *Appl Immunohistochem Mol Morphol* 2011;19:212-7.
- Aperio Technologies, Inc. 510(k) Summary of Substantial Equivalence. Available from: http://www.accessdata.fda.gov/cdrh_docs/pdf7/K071671.pdp. [Last accessed on 2014 Mar 21].
- Olympus Corporation. 510(k) Summary. Virtual Slide System. Olympus VS800 System, VS800HER2 MR Application. Available from: http://www.access.data.fda.gov/cdrh_docs/pdp/7/K111914.pdp. [Last accessed on 2014 Mar 21].
- DAKO Hercept™ Test Package Insert 7th edition. Available from: <http://www.dako.com/us/download.pdf?objectid=121052007>. [Last accessed on 2014 July 01].
- Philips. 510(k) Summary. Philips HER2.neu IHC Digital Manual Read. Available from: http://www.accessdata.fda.gov/cdrh_docs/pdf13/K130021.pdf. [Last accessed on 2014 July 01].
- von Wasielewski R, Krusche CA, Rüschoff J, Fisseler-Eckhoff A, Kreipe H. Implementation of External Quality Assurance Trials for Immunohistochemically Determined Breast Cancer Biomarkers in Germany. *Breast Care (Basel)* 2008;3:128-33.
- Fitzgibbons PL, Murphy DA, Dorfman DM, Roche PC, Tubbs RR, Immunohistochemistry Committee, College of American Pathologists. Interlaboratory comparison of immunohistochemical testing for HER2: Results of the 2004 and 2005 College of American Pathologists HER2 Immunohistochemistry Tissue Microarray Survey. *Arch Pathol Lab Med* 2006;130:1440-5.
- Oluwasola AO, Malaka D, Khramtsov AI, Ikpat OF, Odetunde A, Adeyanju OO, et al. Use of Web-based training for quality improvement between a field immunohistochemistry laboratory in Nigeria and its United States-based partner institution. *Ann Diagn Pathol* 2013;17:526-30.
- Minot DM, Voss J, Rademacher S, Lwin T, Orsulak J, Caron B, 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;137:270-6.
- Gustavson MD, Bourke-Martin B, Reilly D, Cregger M, Williams C, Mayotte J, et al. Standardization of HER2 immunohistochemistry in breast cancer by automated quantitative analysis. *Arch Pathol Lab Med* 2009;133:1413-9.
- 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;28:916-25.
- Turashvili G, Leung S, Turbin D, Montgomery K, Gilks B, West R, 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.
- Rakha EA, Starczynski J, Lee AH, Ellis IO. The updated ASCO/CAP guideline recommendations for HER2 testing in the management of invasive breast cancer: A critical review of their implications for routine practice. *Histopathology* 2014;64:609-15.
- Nassar A, Cohen C, Agersborg SS, Zhou W, Lynch KA, Barker EA, et al. A multisite performance study comparing the reading of immunohistochemical slides on a computer monitor with conventional manual microscopy for estrogen and progesterone receptor analysis. *Am J Clin Pathol* 2011;135:461-7.
- Gudlaugsson E, Skaland I, Janssen EA, Smaalend R, Shao Z, Malpica A, et al. Comparison of the effect of different techniques for measurement of Ki67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer. *Histopathology* 2012;61:1134-44.
- Tsuda H, Kurosuni M, Umemura S, Yamamoto S, Kobayashi T, Osamura RY. HER2 testing on core needle biopsy specimens from primary breast cancers: Interobserver reproducibility and concordance with surgically resected specimens. *BMC Cancer* 2010;10:534.