COLLEGE of AMERICAN PATHOLOGISTS

Validating Whole Slide Imaging Systems for Diagnostic Purposes in Pathology

Guideline from the College of American Pathologists (CAP) Pathology and Laboratory Quality Center.

Effective date: May 1, 2013

CAP WSI Validation Expert Panel Convened June 2010

Liron Pantanowitz, MD, Chair

Bruce Beckwith, MD

Alexis Carter, MD

Lydia Contis, MD

Andrew Evans, MD, PhD

Walter Henricks, MD

Christopher Otis, MD

Anil Parwani, MD, PhD

John Sinard, MD, PhD

Avtar Lal, MD, PhD

Lisa Fatheree, SCT(ASCP) CAP Staff

WSI Validation Key Question

- ***What needs to be done to "validate" a whole slide digital imaging system for diagnostic purposes before it is placed in clinical service?***
- ***Panel addressed:*** ***The intended use, preparation types, number of cases, equipment, personnel, and process.***

CAP

Systematic Review Results

- **Literature search conducted January 2000-2012:**
  - **767 studies met the search term requirements**
  - **27 underwent data extraction for evidence evaluation**
- **Panel met 8 times to develop draft recommendations**
- **Open comment period (July - August 2011):**
  - **132 respondents; 531 comments**
  - **Evidence tables were not completed at that time**
- **Panel met 10 additional times to review feedback, make revisions to draft recommendations, and assess the final data and strength of evidence supporting the 12 recommendations**

CAP

Quality Assessment and Grading of Evidence

- **Strength of evidence:** **level of evidence, quantity, size of the effect, statistical precision and quality assessment (risk of bias) of included studies.**
- **Also taken into account were the study components of consistency, clinical impact, generalizability, and applicability to WSI when determining the strength of evidence score.**

CAP

Definitions of Grading of Recommendations

Definition of Australian National Health and Medical Research Council (NHMRC) to grades of recommendations	
Grade of recommendation	**Description**
A	Body of evidence can be trusted to guide practice
B	Body of evidence can be trusted to guide practice in most situations
C	Body of evidence provides some support for recommendation(s) but care should be taken in its application
D	Body of evidence is weak and recommendation must be applied with caution

Description of Guidance*

*Developed by the CAP Pathology and Laboratory Quality Center

Guidance	Description
Recommendation	For moderate and highest level of evidence (Grade A/B) or where statements are unlikely to change based on further evidence. Note: Can also be in the negative, i.e., 'Recommend Against' or 'Not Recommended'.
Suggestion	For inconclusive, conflicting and/or weak evidence (Grade C) or where statements most likely correct but could be better supported by additional data.
Expert Consensus Opinion	There is a gap, poor evidence (Grade D) or no evidence to support statement but necessary to address the topic. May be qualified with "requires future studies to be conducted".
No recommendation offered	No statement generated for this key question / topic.

CAP

Guideline Statement #1

- **All pathology laboratories implementing WSI technology for clinical diagnostic purposes should carry out their own validation studies.**

  Grade: Expert Opinion

Rationale (#1)

- **Variables between institutions can affect performance.**
- **Manufacturer device validation (ie, verification) alone is insufficient.**
- **Simple guidelines provided for cytology screening devices (which were FDA approved) will not suffice.**

CAP

Guideline Statement #2

- **Validation should be appropriate for and applicable to the intended clinical use and clinical setting of the application in which WSI will be employed. Validation of WSI systems should involve specimen preparation types relevant to intended use (eg, formalin-fixed paraffin-embedded tissue, frozen tissue, immunohistochemical stains, cytology slides, hematology blood smears).**

  Grade: Recommendation, Level A

CAP

Different Outcomes of WSI and Glass Slides with Different Types of Preparation (#2)

Outcomes	Preparations for WSI and Glass Slides					
	H&E		Frozen		Cytology	
	WSI	Glass	WSI	Glass	WSI	Glass
Accuracy of WSI or glass slides	95%	98%	98%	100%	70%	74%
Concordance between WSI and glass slides	84%		94%		100%	
Discordance between WSI and glass slides	16%		6%		0%	
Concordance and minor discordance between WSI and glass slides	97%		97%		100%	

CAP

Guideline Statement #2 Continued

- **Note:** **If a new intended use for WSI is contemplated, and this new use differs materially from the previously validated use, a separate validation for the new use should be performed.**
- **For example:** **A validation study used to support the diagnostic use of digitized slides for routine surgical pathology may not necessarily apply to the use of frozen section digitized slides (eg, with tissue folds, more pale staining, more mounting medium, etc).**

CAP

Guideline Statement #3

- **The validation study should closely emulate the real-world clinical environment in which the technology will be used.**

  Grade: Recommendation, Level A

CAP

Rationale (#3)

- **Goal of validation:**
  - **Conducted in a manner that mimics how WSI will be used in the specific lab's work environment.**
  - **Mimic how the system is to be used after "go live".**
- **For example: If rapid digitization of glass slides is required for clinical use (eg, frozen sections), then timely preparation & reading of WSI should be included in the validation process.**

CAP

Different Outcomes of Whole Slide Imaging (WSI) and Glass Slides with Emulation of Real-World Clinical Environment (#3)

	WSI	Glass Slides
Accuracy of WSI	89%	92%
Concordance between WSI and glass slides	86%	
Discordance between WSI and glass slides	14%	
Concordance and minor discordance between WSI and glass slides	98%	

CAP

Guideline Statement #4

- **The validation study should encompass the entire WSI system.**
- **Note: It is not necessary to validate separately each individual component (eg, computer hardware, monitor, network, scanner) of the system nor the individual steps of the digital imaging process.**

Grade: Recommendation, Level B

CAP

Rationale (#4)

- **WSI system is made up of different components: scanner, hardware, software, network, & viewing monitor (+ pathologist).**
- **Parameters of each of component may impact digital image quality and therefore interpretation.**
- **Imaging process involves several steps including image acquisition, storage, sharing & viewing.**
- **Recommend the *entire* WSI system & imaging process be validated.**
- **All components are important & should not be separated, including technical system (“tool”) & observer (“pathologist”).**

CAP

Different Outcomes of Whole Slide Imaging (WSI) and Glass Slides with Entire WSI System (#4)

	WSI	Glass Slides
Accuracy of WSI	89%	92%
Concordance between WSI and glass slides	83%	
Discordance between WSI and glass slides	17%	
Concordance and minor discordance between WSI and glass slides	98%	

CAP

Guideline Statement #5

- **Revalidation is required whenever a significant change is made to any component of the WSI system.**

  Grade: Expert Opinion

CAP

Rationale (#5)

- **Significant changes to a WSI system may affect the interpretation of digital slides.**
  - **For example: new scanner, major hardware or software upgrade**
- **For major changes the validation process should be repeated:**
  - **With these new changes incorporated in the WSI system**
  - **To demonstrate that it can still be employed for the intended use**
- **Minor changes can be managed through a facilities change management procedure.**

CAP

Guideline Statement #6

- **A pathologist(s) adequately trained to use the WSI system must be involved in the validation process.**

  Grade: Recommendation, Level B

CAP

Rationale (#6)

- **Validation process should include individual(s) who will actually be using the system to make diagnoses.**
- **Published validation studies:**
  - **Average # evaluators = 8 individuals/ study (range, 3 – 26 persons).**
- **Validation team may include other pathology staff**
  - **For example: image technician, histotechnologist, pathologist assistant, IT personnel and/or consultants.**
- **User training is important, but not part of validation.**
  - **Training methods are outside of the scope of this document.**

CAP

Different Outcomes of WSI and Glass Slides with Respect to Training of Pathologists (#6)

Outcome	Training	No Training
	[Mean +/- SD] or Percentage	
Intra-observer agreement of WSI	0.93 ± 0.05	NR
Intra-observer agreement of glass slides	0.93 ± 0.03	NR
Intra-observer agreement between WSI and glass slides	NR	0.71
Inter-observer agreement of WSI	0.82 ± 0.01	0.53 ± 0.11
Inter-observer agreement of glass slides	0.85 ± 0.01	0.59 ± 0.06
Accuracy of WSI	95%	79%
Accuracy of glass slides	99%	81%
Concordance between WSI and glass slides	89%	84%
Discordance between WSI and glass slides	11%	16%
Concordance and minor discordance between WSI and glass slides	98%	98%
Interpretation time of WSI (Min)	4.9 ± 1.6	11.5 ± 2.5

CAP

Guideline Statement #7

- **The validation process should include a sample set of at least 60 cases for one application (eg, H&E stained sections of fixed tissue, frozen sections, cytology, hematology) that reflects the spectrum and complexity of specimen types and diagnoses likely to be encountered during routine operation.**
- *Note: The validation process should include another 20 cases for each additional application (eg, immunohistochemistry, special stains).*

  Grade: Recommendation, Level A

CAP

Rationale #7

- **Validation of WSI systems should:**
  - **Involve specimen preparation types relevant to intended use**
  - ***Not* specific organ systems, diseases, microscopic changes or diagnoses**
- **Important that an adequate sample size be used to allow pathologists to negotiate any technology learning curve.**
- **Literature: Average 92 cases/study (range 10 to 633 cases).**
- *Note: The initial draft for open comment was 100 cases and only received an agreement rate of 73%. The panel made revisions and the evidence supported 60 cases.*

CAP

Different Outcomes of WSI and Glass Slides with Different Number of Cases (#7)

Outcomes	Average Number of Cases		
	20 cases	60 cases	200 cases
Accuracy of WSI	72%	87%	98%*
Accuracy of glass slides	77%	90%	100%
Concordance between WSI and glass slides	75%**	95%	91%
Discordance between WSI and glass slides	25%	5%	9%
Concordance and minor discordance between WSI and glass slides	95%	98%	98%

* P< .001 vs accuracy of 200 cases glass slides

** P= .002 vs concordance of 60 cases & P<.001 vs concordance of 200 cases

CAP

Guideline Statement #8

- **The validation study should establish diagnostic concordance between digital and glass slides for the same observer (ie, intraobserver variability).**

  Grade: Suggestion, Level A

  Due to the conflicting nature of the good quality evidence for accuracy and concordance in both intraobserver and interobserver variability, the statement stands as a Suggestion.

CAP

Different Outcomes of WSI and Glass Slides with Intraobserver and Interobserver Agreement (#8)

Outcome	WSI	Glass Slides
	[Mean +/- SD] or Percentage	
Intra-observer agreement of WSI or glass slides with reference standard	0.93 ± 0.05	0.93 ± 0.03
Intra-observer agreement of WSI and glass slides	0.71	
Inter-observer agreement of WSI or glass slides	0.68 ± 0.06*	0.72 ± 0.04
Concordance between WSI and glass slides	86%	
Concordance and minor discordance between WSI and glass slides	98%	

*P=.005 compared to glass slides

CAP

Rationale #8

- **Baseline intra/inter-observer variability exists even with glass slides.**
- **Aim is to evaluate the technology, not agreement between pathologists.**
  - **eg, Prostate ASAP may have varying pathologist opinions**
- **Therefore, we recommend:**
  - **Measure intraobserver diagnostic reproducibility**

    ie, is the pathologist able to reach the same diagnosis with both modalities?
  - **Don't measure interobserver variability alone**

    ie, their diagnosis compared to other pathologists/ experts/ consensus

- ***Interobserver performance can be performed in conjunction with intraobserver performance if the laboratory desires.***

CAP

Guideline Statement #9

- **Digital and glass slides can be evaluated in random or nonrandom order (as to which is examined first and second) during the validation process.**

  Grade: Recommendation, Level A

CAP

Rationale #9

- **Some believe that digital slides should always be viewed before glass slides - considered the gold standard for making diagnoses.**
- **However, the order of viewing virtual vs. glass slides has been shown not to affect interpretation (*Koch LH et al. Human Pathol 2009; 40:662–7*).**
- **The evidence indicates it can go either way so the laboratory may choose which method to their preference.**
- *Note: The original draft recommendation suggested random order and the panel revised accordingly.*

CAP

Different Outcomes of WSI & Glass Slides with Random or Nonrandom Allocation of Cases (#9)

Outcomes	Allocation of Cases			
	Random		Nonrandom	
	WSI	Glass	WSI	Glass
Accuracy of WSI or glass slides	72%	77%	97%*	99%
Concordance between WSI and glass slides	81%		86%	
Discordance between WSI and glass slides	19%		14%	
Concordance and minor discordance between WSI and glass slides	93%		98%	

*** P< .001 versus glass slide [nonrandom]**

CAP

Guideline Statement #10

- **A washout period of at least 2 weeks should occur between viewing digital and glass slides.**

  Grade: Recommendation, Level B

CAP

Rationale (#10)

- **Washout period: time interval between viewing the same case/slide using a different (glass or digital) modality.**
- **Important to take into consideration:**
  - **Pathologists may recall pathologic images for lengthy periods after reviewing a case.**
  - **With long washout periods a pathologist's experience and/or diagnostic criteria could change over time.**
  - *Note: The original draft recommendation was a 3 week period and it received 68% agreement during open comment. The panel made revisions according to evidence to support 2 weeks minimum.*

CAP

Different Outcomes of WSI and Glass Slides with Different Duration of Washout Periods (#10)

Outcomes	Washout periods for WSI and Glass Slides					
	1 wk		2-3 wk		≥ 6 mo	
	WSI	Glass	WSI	Glass	WSI	Glass
Accuracy of WSI or glass slides	70%	74%	93%	95%	NR	NR
Concordance between WSI and glass slides	NR		87%		95%	
Discordance between WSI and glass slides	NR		13%		5%	
Concordance and minor discordance between WSI and glass slides	NR		95%		100%	

CAP

Guideline Statement #11

- **The validation process should confirm that all of the material present on a glass slide to be scanned is included in the digital image.**

  Grade: Expert Opinion

CAP

Rationale (#11)

- **Accurate digital reproduction of scanned glass slides is required if they are to be used for diagnostic use.**

For example: **What if the WSI device has problems finding small groups of cells or misses cells at the periphery of the slide and therefore doesn't capture any images of them?**

CAP

Guideline Statement #12

- **Documentation should be maintained recording the method, measurements and final approval of validation for the WSI system to be used in the clinical laboratory.**

  Grade: Expert Opinion

CAP

Rationale (#12)

- **Documentation of training should also be recorded.**
- **A statement in the pathology report should also be included that a WSI system was used.**
- **No specific literature addressed documentation.**

CAP

Conclusion

- **Validation of WSI is necessary to ensure that a pathologist using this technique to view digitized glass slides can consistently make the same clinical interpretation as they would from viewing the glass slides using a traditional bright field microscope.**
- **Validation should address both technical and interpretative components, and must be specific for the intended clinical use.**

Pantanowitz L, Sinard JH, Henricks WH, 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. doi: 10.5858/arpa. 2013-0093-CP.

CAP

COLLEGE of AMERICAN PATHOLOGISTS