COLLEGE of AMERICAN PATHOLOGISTS

Supplemental Digital Content | Methodology |
February 2015

Validating Whole Slide Imaging for Diagnostic Purposes in Pathology

Guideline from the Pathology and Laboratory Quality Center

College of American Pathologists | 325 Waukegan Rd. | Northfield, IL 60093 | 800-323-4040 | cap.org

Table of Contents:

Panel Composition

The CAP Pathology and Laboratory Quality Center (the Center) convened an expert panel consisting of members with expertise in digital pathology relevant to whole slide imaging (WSI) and telepathology. Members included practicing United States and Canadian pathologists and CAP staff. CAP approved the appointment of the project, chair and expert panel members.

Conflict of Interest (COI) Policy

Prior to acceptance on the expert panel, potential members completed the CAP conflict of interest (COI) disclosure process, whose policy and form (in effect April 2010) requires 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. Two nominees were excluded from the expert panel after initial COI review in June 2010. Expert panel members' disclosed conflicts are listed in Appendix A of the manuscript. The CAP provided funding for the administration of the project; no industry funds were used in the development of the guideline. Panel members volunteered their time and were not compensated for their involvement.

Methods Used to Produce Guideline

Systematic Literature Review

The charge to the panel was "to recommend validation requirements for whole imaging systems (WSI) used for diagnostic purposes". The central question that the panel addressed was "What should be done to validate a whole slide digital imaging system for diagnostic purposes before it is placed in clinical service?"

A computerized search was conducted during the period from September 28, 2010 to January 23, 2012 in the following electronic databases: OVID MEDLINE, CSA Illumina Conference Papers Index and Google Scholar for articles from January 2000 through January 2012. The search utilized the following terms:

- Whole slide imaging OR Virtual or Digital microscopy OR Digital pathology OR Teleconsultation OR Telemicroscopy – AND
- Validation

Alternate terms 'digitized slide' and 'whole slide scanner' were also used. Reference lists from identified articles were scrutinized for articles not identified in the above search.

Eligible Study Designs

The search included all types of study design. In addition to articles, the search identified published abstracts presented at various conferences including international. The search was not limited to the English language, and one Russian article was included for the full text review.

Inclusion Criteria

Studies were selected for full text review based upon the following criteria:
(1) the study referred to whole slide imaging, and
(2) the study pertained to clinical use or investigative research.
All clinical fields (e.g., pathology, veterinary, etc.) were allowed.

Exclusion Criteria

Publications involving static and robotic digital imaging, purely technical components, only educational applications, and image analysis were excluded.

Outcomes of Interest

Validation requirements addressed included intended use, preparation types, number of cases, equipment, personnel, and process. Analysis of the evidence included accuracy, concordance, sensitivity, specificity, inter- and intra-observer variability, and average diagnostic certainty. Interpretation and scanning time were also reviewed but not included in the final recommendations or the manuscript.

Environmental Scan

In the United States, the Food and Drug Administration (FDA) convened a Hematology and Pathology Devices Panel hearing in October 2009 that focused on how best to regulate WSI systems that are to be used for primary diagnosis in surgical pathology.[1] http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/HematologyandPathologyDevicesPanel/UCM187186.pdf. After the October 2011 Pathology Visions Meeting, CAP Today published an article summarizing the FDA stance at that time. http://www.cap.org/apps/cap.portal?_nfpb=true&cntvwrPtlt_actionOverride=%2Fportlets%2FcontentViewer%2Fshow&_windowLabel=cntvwrPtlt&cntvwrPtlt{actionForm.contentReference}=cap_today%2F0112%2F0112a_regulators.html&_state=maximized&_pageLabel=cntvwr (accessed October 2012).[2]

Quality Assessment and Grading of the Included Evidence

(Hierarchy of evidence, body of evidence, definitions of grades of recommendations)

The literature review was performed in duplicate by two members of the expert panel. A third reviewer was involved if they were not able to reach consensus. A contracted methodologist and CAP staff performed final data extraction. Each study was assessed for strength of evidence, which consisted of level of evidence, quantity, size of the effect, statistical precision and, quality (risk of bias). The quality assessment of the studies was performed by using the *Whiting et al 2003* instrument.[3] The other components of evidence such as consistency, clinical impact, generalizability, and applicability to digital pathology were also considered when determining the strength of evidence (Table 1). An overall grade for a recommendation was obtained by considering the component scores of individual items.[4]

For ***strength*** of the evidence, we considered the level of evidence, its quantity, size of the overall effect, statistical precision, and quality of included studies. The ***level*** of evidence was based on the study design as follows: Level I was evidence from systematic reviews of appropriate level II studies; level II was evidence from good quality diagnostic studies or randomized controlled trials; level III was evidence from low quality comparative diagnostic studies; level IV was evidence from diagnostic studies without a reference standard. Level I and II evidence was considered most appropriate to answer the clinical question put to this panel. The ***quantity*** of evidence refers to the number of studies and number of patients/cases included for each outcome in the recommendation. The ***size of the effect*** refers to the overall effect and its statistical precision. It was measured as weighted mean difference or risk ratio and confidence intervals. The ***quality of studies*** reflected how well the studies were designed to eliminate bias, including how the subjects, cases or tests were selected, allocated to groups (study or test and control or standard), managed, followed up, and analyzed. The methodological quality of diagnostic study was critically appraised using *Whiting et al 2003* checklist.[3] All these components of the evidence base were considered while allocating an overall score to strength of evidence.[4]

For consistency, we assessed both the clinical and statistical heterogeneity among the studies. The clinical heterogeneity was the variability regarding patients, disease state, and type of test to diagnose a condition and

its comparator, and outcome measured. In the presence of marked clinical heterogeneity the studies were not meta-analyzed nor subgroup meta-analysis was performed. The statistical heterogeneity was assessed by performing a meta-analysis and was measured as I^2 and P value.[4, 5]

For clinical impact, we assessed the potential benefits of test or intervention to the population and the relevance of evidence to the clinical question of the recommendation. In addition, we also considered the size of the effect, its statistical precision, and the relevance of effect of a test or intervention to the patients compared with other management options if available. The clinical impact could vary from very large to slight clinical impact.[4, 5]

For generalizability, we observed how well the subjects and settings of the included studies matched those of the recommendation. Population parameters such as gender, age, ethnicity, and baseline risk, and the level of care were considered. If the population studied in the body of evidence was the same as the target population for the guideline, it was scored as excellent. On the other hand, if these populations were substantially different it was scored poor.[4, 5]

For applicability, we considered how well the entire evidence favoring the recommendations was relevant to the United States and international populations. Evidence which was directly applicable to the United States and international healthcare scored as excellent, whereas that which was not scored as poor.[4, 5]

Table 1: Body of Evidence Matrix Component[4]

	A	B	C	D
	Excellent	**Good**	**Satisfactory**	**Poor**
Evidence base	several level I or level II studies with low risk of bias	one or two level II studies with low risk of bias or a systematic review/multiple level III studies with low risk of bias	level III studies with low risk of bias, or level I or II studies with moderate risk of bias	level IV studies, or level I to III studies with high risk of bias
Consistency	all studies consistent	most studies consistent and inconsistency may be explained	some inconsistency reflecting genuine uncertainty around clinical question	evidence is inconsistent
Clinical impact	very large	substantial	moderate	slight or restricted
Generalizability	population/s studied in body of evidence are the same as the target population for the guideline	population/s studied in the body of evidence are similar to the target population for the guideline	population/s studied in body of evidence differ from target population for guideline but it is clinically sensible to apply this evidence to target population	population/s studied in body of evidence differ from target population and hard to judge whether it is sensible to generalize to target population
Applicability	directly applicable to United States (US) and international healthcare context	applicable to US and international healthcare context with few caveats	probably applicable to US and international healthcare context with some caveats	not applicable to US or international healthcare context

The overall grade of each recommendation was obtained by rating all components of the evidence. The overall grade indicates the strength of the body of evidence to assist the users of clinical practice guidelines in making appropriate and informed clinical judgments (Table 2). Grade A or B evidence supports “recommendations”, the term we use for guidance based on a body of evidence that can be trusted to guide clinical practice in all or most situations. Grade C evidence is insufficient to support a “recommendation”; instead we use the term “suggestion”, for which care should be taken in application. “Suggestions” may also reflect guidance in cases where the evidence is conflicting or inconclusive. Grade D evidence is weak and does not provide support for either “recommendations” or “suggestions”. However, the guideline authors may choose to provide guidance in the form of an “expert consensus opinion” where they believe that guidance will result in improved patient care, even in cases where the evidence is low or lacking (Table 3). In this guideline, guidance includes recommendations, suggestions and expert consensus opinion; there were no instances of “no recommendation offered”.

Table 2: Definition of Grades of Recommendations[4]

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

Table 3: Description of Guidance*

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.

*Developed by the CAP Pathology and Laboratory Quality Center

Revision Dates

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

Outcomes

CAP Expert Panel Literature Review and Analysis

Initially, the chair sent a Zoomerang study with 58 potential digital pathology validation statements to all expert panel members and instructed them to respond with Agree, Disagree, or Don't Know and provide comment. Results included 7 statements with 100% agreement, 9 with 88% agreement, 16 with 75% agreement, 10 with 62% agreement, 8 with 50% agreement, 4 with 38% agreement, 3 with 25% agreement, and 1 statement with a 0% agreement. The expert panel proceeded to review all statements and discuss those with an agreement rate of 62% or less. During discussion, variable interpretation of the statement was found to be the largest cause of disagreement amongst members. Resolution was obtained by majority consensus and many statements were eliminated from recommendation consideration, considered duplicate statements or reduced to comments of interest to address in the manuscript.

The expert panel met in a face-to-face meeting September 2010; additional work was completed through 18 teleconference webinars, collaboration site access (Oracle WebCenter Spaces v11.1.1.2.0) and electronic mail. The purpose of the panel meeting was to refine the scope of the document and address the most discordant Zoomerang digital pathology validation statements amongst panel members.

An open comment period was held from July 22, 2011 through August 21, 2011. Thirteen statements (representing potential recommendations) with brief background information and an open ended question were posted online on the CAP website. An announcement was sent to the following societies:

- College of American Pathologists (CAP)
- Association of Directors of Anatomic and Surgical Pathology (ADASP)
- American Society of Clinical Pathology (ASCP)

- Association for Pathology Informatics (API)
- Digital Pathology Association (DPA)
- International Academy of Digital Pathology (IADP)
- Association for Pathology Chairs (APC)
- Canadian Association of Pathologists (CAP-APC)
- Food and Drug Administration (FDA)

The website received 531 comments in total (Agree, Disagree, and Comment). Each expert panel member was assigned 1-2 statements for which to review all comments received and provide an overall summary to the rest of the panel. Following their review the panel members determined whether to maintain the original statement/recommendation as is, revise it with minor language change, or consider a major statement/recommendation change. Based upon the fact that most statements achieved over 80% agreement with the original recommendation, the expert panel elected to make minor modifications to the statements for clarification and/or explain any pertinent issues in further detail within the manuscript.

Only two statements did not achieve 80% agreement and the expert panel accordingly made major revisions of these recommendations. Seven statements were revised with only minor language changes and four statements were maintained with the original language. Additional revisions were made by the panel after the quality of evidence was assessed. Resolution of major and minor changes was obtained by majority consensus of the panel.

767 studies met the search term requirements (Figure 1). For title/abstract review, each study underwent an inclusion-exclusion, dual independent review conducted by staff, chair and a third member referee when staff/chair review did not achieve unanimous agreement. The initial title/abstract review eliminated 655 studies. Dual independent expert panel members and staff reviewed the remaining 112 studies in full.

To include the study for grading by the methodologist, composite scoring by both reviewers had to be a score of four or above (Figure 2). The expert panel members unanimously eliminated 31 studies and the chair eliminated an additional 54 studies for total of 85 exclusions in full text review. Twenty seven studies received a strong enough score to be considered for data extraction and review by the contracted methodologist. After data extraction verification by CAP staff, 23 studies were included in the final evidence. Any excluded article was available as discussion or background references.

The expert panel performed preliminary data extraction in the following areas: year of publication, country of origin, publication type, application of study, subspecialty of study, number of pathologists (or individuals), numbers of cases, validation method, reported concordance and outcome measurement. All members of the expert panel participated in the draft manuscript.

An independent review panel (IRP) was assembled to review the guideline and recommend approval to the CAP Transformation Program Office Steering Committee, which had final approval authority. The IRP was masked to the expert panel and vetted through the COI process. Because of the nature of the content, input from industry was considered. The Executive Committee of the DPA was sent a courtesy copy during the final review process.

The CAP Pathology and Laboratory Quality Center Disclaimer

The CAP developed the Pathology and Laboratory Quality Center as a forum to create and maintain evidence based practice guidelines and consensus statements. Practice guidelines and consensus statements reflect the best available evidence and expert consensus supported in practice. They are intended to assist physicians and patients in clinical decision-making and to identify questions and settings for further research. With the rapid flow of scientific information, new evidence may emerge between the time a practice guideline or consensus statement is developed and when it is published or read. Guidelines and statements are not continually updated and may not reflect the most recent evidence. Guidelines and statements address only the topics specifically identified therein and are not applicable to other interventions, diseases, or stages of diseases. Furthermore, guidelines and consensus statements cannot account for individual variation among patients and cannot be considered inclusive of all proper methods of care or exclusive of other treatments. It is the responsibility of the treating physician or other health care provider, relying on independent experience and knowledge, to determine the best course of treatment for the patient. Accordingly, adherence to any practice guideline or consensus statement is voluntary, with the ultimate determination regarding its application to be made by the

physician in light of each patient's individual circumstances and preferences. CAP makes no warranty, express or implied, regarding guidelines and statements and specifically excludes any warranties of merchantability and fitness for a particular use or purpose. CAP assumes no responsibility for any injury or damage to persons or property arising out of or related to any use of this statement or for any errors or omissions.

Guideline Statements

Summary of Evidence Tables

1: Expert Consensus Opinion

All pathology laboratories considering the implementation of WSI technology for clinical diagnostic purposes should carry out their own validation studies.

Evidence Review: There was no published data directly linked to the statement.

2: Recommendation

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 (e.g. formalin-fixed paraffin-embedded tissue, frozen tissue, immunohistochemical stains, cytology slides, hematology blood smears).
Grade: Level A

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.

Evidence Review: For different types of preparations (such as Hematoxylin and Eosin (H&E) stains of fixed tissue, frozen tissue, cytology and immunostains), our meta-analysis showed no significant difference in the accuracy between WSI and glass slides when compared with the reference standard. There was good concordance between the WSI and glass slides.
Evidence base: A
Consistency: A
Clinical impact: A
Generalizability: A
Applicability: A
Overall Grade: A

Table 4: Different outcomes of WSI and Glass Slides with different types of preparation

Outcomes	**Preparations for WSI and Glass Slides**					
	H&E		**Frozen**		**Cytology**	
	WSI	**Glass**	**WSI**	**Glass**	**WSI**	**Glass**
Accuracy of WSI or glass slides[6-13]	95%	98%	98%	100%	70%	74%
Concordance between WSI and glass slides[9, 10, 12-27]	84%		94%		100%	
Discordance between WSI and glass slides[9, 10, 12-27]	16%		6%		0%	
Concordance and minor discordance between WSI and glass slides[9, 10, 12, 14-21, 23-27]	97%		97%		100%	

3: Recommendation

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

Evidence Review: Our meta-analysis of cases in a real-world clinical environment (e.g. routine slides employed) showed no significant difference in the accuracy between WSI or glass slides when compared with the reference standard. There was good concordance between diagnoses made using WSI and glass slides.
Evidence base: A
Consistency: A
Clinical impact: A
Generalizability: A
Applicability: A
Overall Grade: A

Table 5: Different outcomes of WSI and Glass Slides with emulation of the real-world clinical environment

	WSI	Glass Slides
Accuracy of WSI[6-12]	89%	92%
Concordance between WSI and glass slides[9, 10, 12, 14-27]	86%	
Discordance between WSI and glass slides[9, 10, 12, 14-27]	14%	
Concordance and minor discordance between WSI and glass slides[9, 10, 12, 14-21, 23, 24, 26, 27]	98%	

4: Recommendation
The validation study should encompass the entire WSI system.

Note: It is not necessary to validate separately each individual component (e.g., computer hardware, monitor, network, scanner) of the system nor the individual steps of the digital imaging process.
Evidence Review: Some studies reported validation of different components of the WSI system (e.g. scanning device, transferring images onto the web/Internet, Internet connectivity, configuration of computers, monitors and software used). Our meta-analysis of those studies showed no significant difference in the accuracy of WSI and glass slides, when compared with the reference standard. There was good concordance between WSI and glass slides.
Evidence base: B
Consistency: B
Clinical impact: A
Generalizability: A
Applicability: A
Overall Grade: B

Table 6: Different outcomes of WSI and Glass Slides with the entire WSI system

	WSI	Glass Slides
Accuracy of WSI[6, 7, 9-13]	89%	92%
Concordance between WSI and glass slides[9, 10, 13-16, 18, 20-22]	83%	
Discordance between WSI and glass slides[9, 10, 13-16, 18, 20-22]	17%	
Concordance and minor discordance between WSI and glass slides[9, 10, 13-16, 18, 20]	98%	

5: Expert Consensus Opinion
Re-validation is required whenever a significant change is made to any component of the WSI system.

Evidence Review: There was no published data on re-validation to analyze.

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

Evidence Review: In some studies, pathologists were appropriately trained on using the WSI system. However, in other studies, training on using the WSI system was either not imparted or it was not reported.. Our analysis showed that when training on using WSI system was imparted to pathologists, there tended to be greater accuracy of WSI, better concordance (between WSI and glass slides), and a shorter interpretation time compared to no training. No study directly evaluated the effect of training and no training on any outcome of WSI system.
Evidence base: B
Consistency: B
Clinical impact: A
Generalizability: B
Applicability: A
Overall Grade: B

Table 7: Different outcomes of WSI and Glass Slides with training of pathologists

	Training	**No Training**
	[Mean +/- SD] or Percentage	
Intra-observer agreement of WSI[12]	0.93 ± 0.05	NR
Intra-observer agreement of glass slides[12]	0.93 ± 0.03	NR
Intra-observer agreement between WSI and glass slides[12]	NR	0.71
Inter-observer agreement of WSI[12, 28]	0.82 ± 0.01	0.53 ± 0.11
Inter-observer agreement of glass slides[12, 28]	0.85 ± 0.01	0.59 ± 0.06
Accuracy of WSI[6, 9, 12, 13]	95%	79%
Accuracy of glass slides[6, 9, 12, 13]	99%	81%
Concordance between WSI and glass slides[9, 10, 13-18, 20, 22, 24, 26]	89%	84%
Discordance between WSI and glass slides[9, 10, 13-18, 20, 22, 24, 26]	11%	16%
Concordance and minor discordance between WSI and glass slides[9, 10, 13-18, 20, 26]	98%	98%
Interpretation time of WSI (Min)[6, 7, 13, 21, 24, 26]	4.9 ± 1.6	11.5 ± 2.5

NR = Not reported

7: *Recommendation*

The validation process should include a sample set of at least 60 cases for one application (e.g. 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 (e.g. immunohistochemistry, special stains).*

Evidence Review: Different studies reported using a different number of cases in evaluation.
An average of 20 cases (range 10 to 46) showed a tendency toward less accurate diagnoses made by WSI compared to glass slides. The concordance (between WSI and glass slides) was significantly less with an average of 20 cases compared to an average of 60 cases (range 52 to 90) (P< 0.002) and 200 cases (range 100 to 633) (P<0.001).
There was no significant difference in the accuracy (between diagnoses made using WSI and glass slides) with an average of 60 cases; however, accuracy was significantly lower (P< 0.0004) with WSI compared to glass slides with 200 cases. The concordance (between WSI and glass slides) was good for 60 or 200 cases.
The use of at least 60 cases is recommended as it tends to result in better accuracy and concordance than an average of 20 cases and almost similar accuracy and concordance to an average of 200 cases.
Evidence base: A
Consistency: A
Clinical impact: A
Generalizability: A
Applicability: A
Overall Grade: A

Table 8: Different outcomes of WSI and Glass Slides with different number of cases

Outcomes	**Average Number of Cases**		
	20 cases	**60 cases**	**200 cases**
Accuracy of WSI[6-13, 25]	72%	87%	98%*
Accuracy of glass slides[6-13, 25]	77%	90%	100%
Concordance between WSI and glass slides[9, 10, 13-27]	75%**	95%	91%
Discordance between WSI and glass slides[9, 10, 13-27]	25%	5%	9%
Concordance and minor discordance between WSI and glass slides[9, 10, 13-18, 20, 21, 23, 26, 27]	95%	98%	98%

*P< 0.0004 vs accuracy of 200 cases glass slides; ** P< 0.002 vs concordance of 60 cases and P<0.001 vs concordance of 200 cases

8: Suggestion
The validation study should establish diagnostic concordance between digital and glass slides for the same observer (i.e., intraobserver variability).

Evidence Review: There was good and similar intra-observer agreement for the WSI and glass slides when compared with the reference standard. The inter-observer agreement was significantly less (P < 0.005) for WSI comparisons to glass slides. There was good concordance between the WSI and glass slides for both intra- and inter-observer agreement. Due to the conflicting nature of the good quality evidence for agreement and concordance, the statement stands as a Suggestion.
Evidence base: B
Consistency: B
Clinical impact: A
Generalizability: A
Applicability: A
Overall Grade: A

Table 9: Different outcomes of WSI and Glass Slides with intra- and inter-observer agreement

	WSI	Glass Slides
	[Mean +/- SD] or Percentage	
Intra-observer agreement of WSI or glass slides with reference standard[6-13]	0.93 ± 0.05	0.93 ± 0.03
Intra-observer agreement of WSI and glass slides[6-13]	0.71	
Inter-observer agreement of WSI or glass slides[6-13]	0.68 ± 0.06*	0.72 ± 0.04
Concordance between WSI and glass slides[9, 10, 13-20, 23]	86%	
Concordance and minor discordance between WSI and glass slides[9, 10, 13-18, 20, 23]	98%	

*P<0.005 compared to glass slides

9: Recommendation

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

Evidence Review: Few studies evaluated digital and glass slides in random order. Our meta-analysis showed no significant difference in the accuracy between WSI and glass slides in studies following random allocation of WSI and glass slides, but there was significantly lower accuracy (P = 0.0003)of WSI compared to glass slides in studies following non-random allocation.

There was no marked difference in the concordance in studies of random allocation compared to studies of no random allocation.

Evidence base: A
Consistency: A
Clinical impact: A
Generalizability: A
Applicability: A
Overall Grade: A

Table 10: Different outcomes of WSI and Glass Slides with random or non-random allocation of cases

Outcomes	**Allocation of Cases**			
	Random		**Non-Random**	
	WSI	**Glass**	**WSI**	**Glass**
Accuracy of WSI or glass slides[7, 10, 12]	72%	77%	97%*	99%
Concordance between WSI and glass slides[10, 14, 15, 20]	81%		86%	
Discordance between WSI and glass slides[10, 14, 15, 20]	19%		14%	
Concordance and minor discordance between WSI and glass slides[10, 14, 15, 20]	93%		98%	

*P = 0.0003 vs glass slide [non-random]

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

Evidence Review: Few studies reported washout periods while examining WSI and glass slides; these papers included washouts of 1 week (Evered A 2011), 2 weeks (Molnar B 2003), and approximately 3 weeks (Jukic DM 2011, Nielsen PS 2010,). No study compared the outcomes with different washout periods.
A washout period of at least 2 weeks showed good accuracy and concordance between WSI and glass slides. Due to the limited amount of published data, the effect of other washout periods on the accuracy and concordance between WSI and glass slides remains unclear at this time.
Evidence base: B
Consistency: B
Clinical impact: A
Generalizability: A
Applicability: A
Overall Grade: B

Table 11: Different outcomes of WSI and Glass Slides with different duration of washout periods

Outcomes	Washout periods for WSI and Glass Slides					
	1 Week		2-3 Weeks		≥ 6 months	
	WSI	Glass	WSI	Glass	WSI	Glass
Accuracy of WSI or glass slides[12, 28]	70%	74%	93%	95%	NR	NR
Concordance between WSI and glass slides[9, 14-16, 18, 20, 23]	NR		87%		95%	
Discordance between WSI and glass slides[9, 14-16, 18, 20, 23]	NR		13%		5%	
Concordance and minor discordance between WSI and glass slides[9, 14-16, 18, 20, 23]	NR		95%		100%	

NR = Not reported

11: Expert Consensus Opinion

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

Evidence Review: There was no published data specifically addressing the presence/absence of all material to analyze.

12: Expert Consensus Opinion

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

Evidence Review: There was no published peer-reviewed data on documentation to analyze.

Figure 1: Literature Review Results

Adapted with permission from Moher et al.[29]

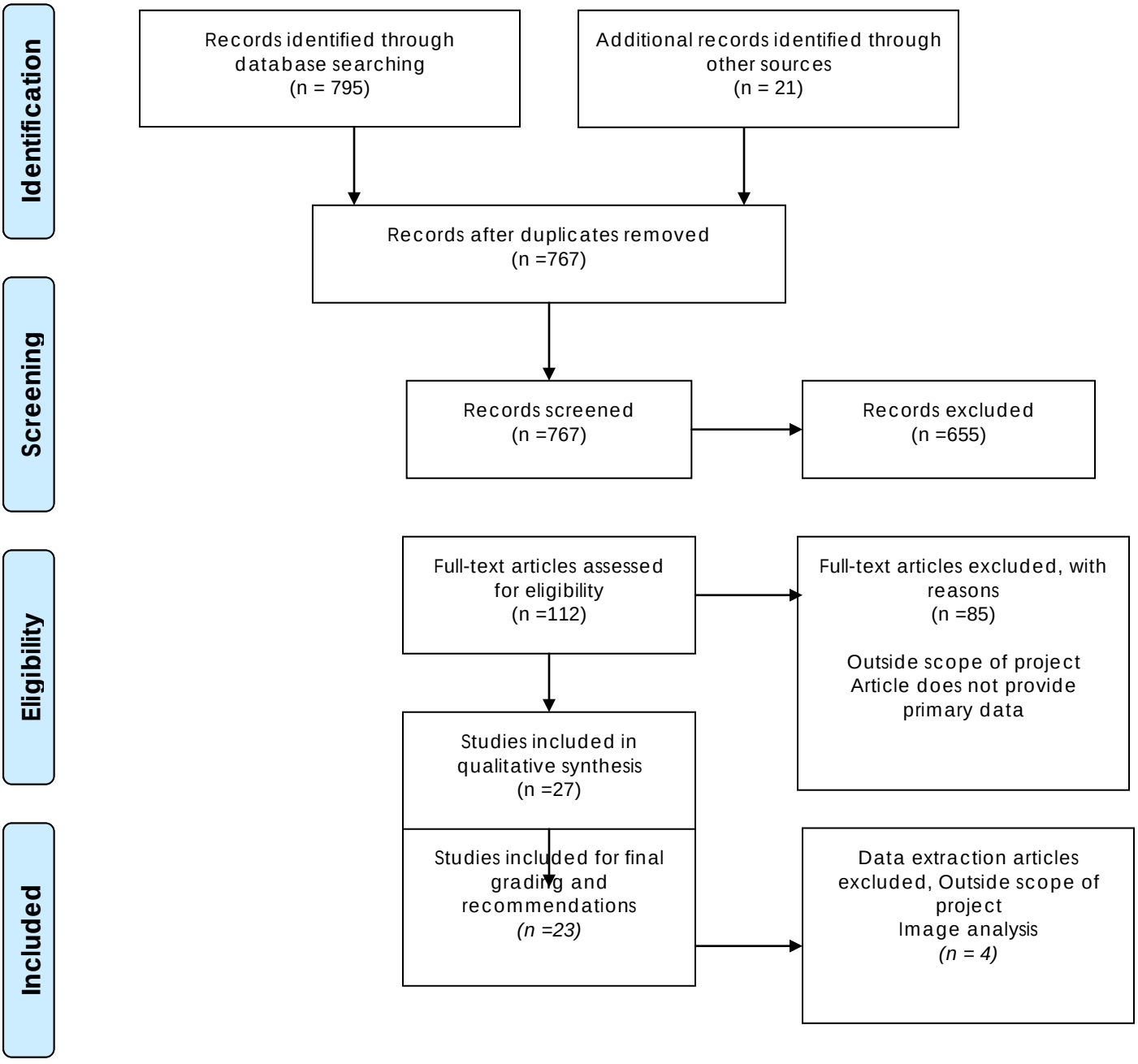
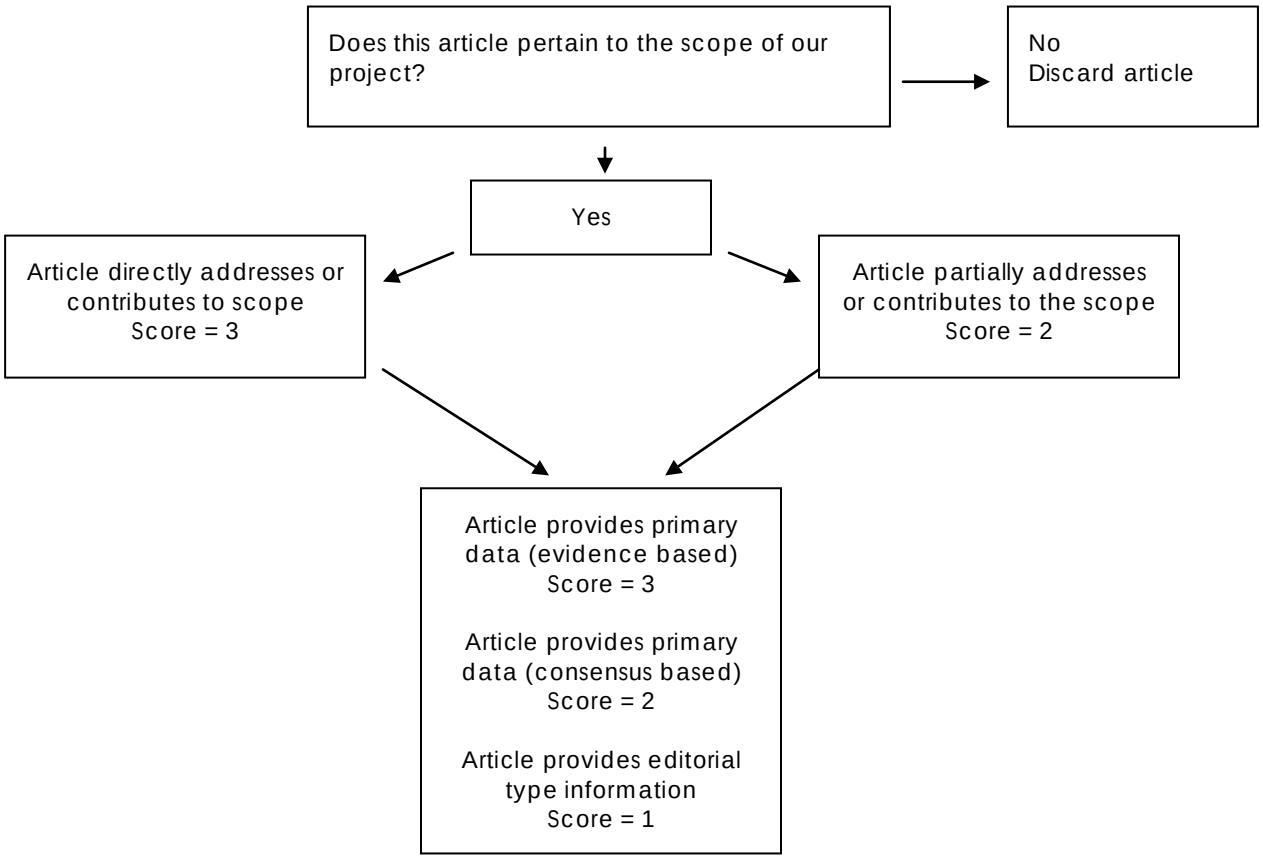


Figure 2: Full Text Criteria



References:

1. Lusky K. Where will FDA land on whole-slide digital? *CAP Today*. 2009;23(12):1.
2. Titus K. Regulators scanning the digital scanners. *CAP Today*. 2012;26(1):56-62.
3. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
4. Hillier S, Grimmer-Somers K, Merlin T, et al. FORM: an Australian method for formulating and grading recommendations in evidence-based clinical guidelines. *BMC Med Res Methodol*. 2011;11:23.
5. Merlin T, Weston A, Tooher R. Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence'. *BMC Med Res Methodol*. 2009;9:34.
6. Evans AJ, Chetty R, Clarke BA, et al. Primary frozen section diagnosis by robotic microscopy and virtual slide telepathology: the University Health Network experience. *Hum Pathol*. 2009;40(8):1070-1081.
7. Evered A, Dudding N. Accuracy and perceptions of virtual microscopy compared with glass slide microscopy in cervical cytology. *Cytopathology*. 2011;22(2):82-87.
8. Furness P. A randomized controlled trial of the diagnostic accuracy of internet-based telepathology compared with conventional microscopy. *Histopathology*. 2007;50(2):266-273.
9. Ho J, Parwani AV, Jukic DM, Yagi Y, Anthony L, Gilbertson JR. Use of whole slide imaging in surgical pathology quality assurance: design and pilot validation studies. *Hum Pathol*. 2006;37(3):322-331.
10. Molnar B, Berczi L, Diczhazy C, et al. Digital slide and virtual microscopy based routine and telepathology evaluation of routine gastrointestinal biopsy specimens. *J Clin Pathol*. 2003;56(6):433-438.
11. Mooney E, Hood AF, Lampros J, Kempf W, Jemec GB. Comparative diagnostic accuracy in virtual dermatopathology. *Skin Res Technol*. 2011;17(2):251-255.
12. Nielsen PS, Lindebjerg J, Rasmussen J, Starklint H, Waldstrom M, Nielsen B. Virtual microscopy: an evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors. *Hum Pathol*. 2010;41(12):1770-1776.
13. Velez N, Jukic D, Ho J. Evaluation of 2 whole-slide imaging applications in dermatopathology. *Hum Pathol*. 2008;39(9):1341-1349.
14. Al-Janabi S, Huisman A, Vink A, et al. Whole slide images for primary diagnostics of gastrointestinal tract pathology: a feasibility study. *Hum Pathol*. 2012;43(5):702-707.
15. Al-Janabi S, Huisman A, Vink A, et al. Whole slide images for primary diagnostics in dermatopathology: a feasibility study. *J Clin Pathol*. 2012;65(2):152-158.
16. Costello SS, Johnston DJ, Dervan PA, O'Shea DG. Development and evaluation of the virtual pathology slide: a new tool in telepathology. *J Med Internet Res*. 2003;5(2):e11.
17. Fallon MA, Wilbur DC, Prasad M. Ovarian frozen section diagnosis: use of whole-slide imaging shows excellent correlation between virtual slide and original interpretations in a large series of cases. *Arch Pathol Lab Med*. 2010;134(7):1020-1023.
18. Gilbertson JR, Ho J, Anthony L, Jukic DM, Yagi Y, Parwani AV. Primary histologic diagnosis using automated whole slide imaging: a validation study. *BMC Clin Pathol*. 2006;6:4.
19. Henderson J, Scott K, Carpenter J, Chiang S, Weinstein R. Diagnostic accuracy of quality assurance surgical pathology specimens using virtual microscopy. *Arch Pathol Lab Med*. 2007;131(9):1429-1430.
20. Jukic DM, Drogowski LM, Martina J, Parwani AV. Clinical examination and validation of primary diagnosis in anatomic pathology using whole slide digital images. *Arch Pathol Lab Med*. 2011;135(3):372-378.
21. Lopez AM, Graham AR, Barker GP, et al. Virtual slide telepathology enables an innovative telehealth rapid breast care clinic. *Semin Diagn Pathol*. 2009;26(4):177-186.
22. Massone C, Soyer HP, Lozzi GP, et al. Feasibility and diagnostic agreement in teledermatopathology using a virtual slide system. *Hum Pathol*. 2007;38(4):546-554.
23. Pagni F, Bono F, Di Bella C, Faravelli A, Cappellini A. Virtual surgical pathology in underdeveloped countries: The Zambia Project. *Arch Pathol Lab Med*. 2011;135(2):215-219.
24. Slodkowska J, Chyczewski L, Wojciechowski M. Virtual slides: application in pulmonary pathology consultations. *Folia Histochem Cytobiol*. 2008;46(1):121-124.
25. Wilbur DC, Madi K, Colvin RB, et al. Whole-slide imaging digital pathology as a platform for teleconsultation: a pilot study using paired subspecialist correlations. *Arch Pathol Lab Med*. 2009;133(12):1949-1953.
26. Ramey J, Fung KM, Hassell LA. Use of mobile high-resolution device for remote frozen section evaluation of whole slide images. *J Pathol Inform*. 2011;2:41.

27. Slodkowska J, Pankowski J, Siemiatkowska K, Chyczewski L. Use of the virtual slide and the dynamic real-time telepathology systems for a consultation and the frozen section intra-operative diagnosis in thoracic/pulmonary pathology. *Folia Histochem Cytobiol*. 2009;47(4):679-684.
28. Fine JL, Grzybicki DM, Silowash R, et al. Evaluation of whole slide image immunohistochemistry interpretation in challenging prostate needle biopsies. *Hum Pathol*. 2008;39(4):564-572.
29. Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.