UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ODONTOLOGIA DE PIRACICABA

ANNA LUIZA DAMACENO ARAUJO

VALIDAÇÃO DA MICROSCOPIA DIGITAL NO DIAGNÓSTICO HISTOPATOLÓGICO DE DOENÇAS BUCAIS

VALIDATION OF DIGITAL MICROSCOPY IN THE HISTOPATHOLOGICAL DIAGNOSES OF ORAL DISEASES

Piracicaba

2018

ANNA LUIZA DAMACENO ARAUJO

VALIDAÇÃO DA MICROSCOPIA DIGITAL NO DIAGNÓSTICO HISTOPATOLÓGICO DE DOENÇAS BUCAIS

VALIDATION OF DIGITAL MICROSCOPY IN THE HISTOPATHOLOGICAL DIAGNOSES OF ORAL DISEASES

Dissertação apresentada à Faculdade de Odontologia de Piracicaba da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Estomatopatologia, na Área de Estomatologia.

Dissertation presented to the Piracicaba Dental School of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Stomatopathology, in Stomatology area.

Orientador: Prof. Dr. Alan Roger dos Santos Silva

Este exemplar corresponde a versão final da dissertação defendida pela aluna Anna Luiza Damaceno Araujo e orientada pelo Prof. Dr. Alan Roger dos Santos Silva.

Piracicaba

2018

Agência(s) de fomento e nº(s) de processo(s): CAPES, 33003033009P4
ORCID: https://orcid.org/0000-0002-3725-8051

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Odontologia de Piracicaba
Marilene Girello - CRB 8/6159

Ar15v

Araujo, Anna Luiza Damaceno, 1990-
Validação da microscopia digital no diagnóstico histopatológico de doenças bucais / Anna Luiza Damaceno Araujo. – Piracicaba, SP : [s.n.], 2018.

Orientador: Alan Roger dos Santos Silva.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Odontologia de Piracicaba.

1. Estudos de validação. 2. Microscopia. 3. Revisão. 4. Boca - Doenças. I. Santos-Silva, Alan Roger, 1981-. II. Universidade Estadual de Campinas. Faculdade de Odontologia de Piracicaba. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Validation of digital microscopy in the histopathological diagnoses of oral diseases
Palavras-chave em inglês:
Validation studies
Microscopy
Review
Mouth - Diseases
Área de concentração: Estomatologia
Titulação: Mestra em Estomatopatologia
Banca examinadora:
Alan Roger dos Santos Silva [Orientador]
Mário José Romañach Gonzalez Sobrinho
Pablo Agustin Vargas
Data de defesa: 25-07-2018
Programa de Pós-Graduação: Estomatopatologia

UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Odontologia de Piracicaba

A Comissão Julgadora dos trabalhos de Defesa de Dissertação de Mestrado, em sessão pública realizada em 25 de Julho de 2018, considerou a candidata ANNA LUIZA DAMACENO ARAUJO aprovada.

PROF. DR. ALAN ROGER DOS SANTOS SILVA

PROF. DR. MÁRIO JOSÉ ROMAÑACH GONZALEZ SOBRINHO

PROF. DR. PABLO AGUSTIN VARGAS

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno.

AGRADECIMENTOS

A Deus, que com sua magnitude me cedeu força e me reergueu nos momentos difíceis.

À Universidade Estadual de Campinas, na pessoa do Magnífico Reitor, Prof. Dr. Marcelo Knobel.

À Faculdade de Odontologia de Piracicaba, na pessoa de seu Diretor, Prof. Dr. Guilherme Elias Pessanha Henriques e seu Diretor Associado, Prof. Dr. Francisco Haiter Neto.

À Profa. Dra. Cínthia Pereira Machado Tabchoury, Coordenadora Geral da Pós-Graduação da Faculdade de Odontologia de Piracicaba.

Ao Coordenador do Programa de Pós-Graduação em Estomatopatologia, Prof. Dr. Marcio Ajudarte Lopes.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/PROEX) pela concessão da bolsa para a realização dessa dissertação de Mestrado.

Ao meu orientador, Prof. Dr. Alan Roger dos Santos Silva, por toda a ajuda desde o primeiro dia que cheguei à Piracicaba. Seu profissionalismo e orientação me inspiram a crescer cada vez mais nessa caminhada profissional. Agradeço pelas oportunidades, pela paciência e por sempre valorizar e levar em consideração meus posicionamentos e opiniões. Confio bastante na sua conduta, como orientador e como pessoa, e espero poder evoluir ainda mais com seus ensinamentos. Muito obrigada por tudo, sempre.

A Gleyson Kleber do Amaral Silva e Christian Camilo Madrid Troconis, que foram peças fundamentais na execução deste trabalho. Obrigada pelo tempo dispendido, pela disposição e paciência ao compartilhar tanto conhecimento comigo. Sem vocês não teríamos conquistado tanto.

Aos professores Pablo Agustin Vargas e Oslei Paes de Almeida, por terem participado ativamente desta pesquisa e contribuído em tantos pontos importantes que aperfeiçoaram a qualidade deste trabalho.

A todos os amigos que a pós-graduação me concedeu, obrigada pelos cafés, conselhos e risadas. É muito importante ter uma rede de apoio em todos os momentos das nossas vidas e, encontrar pessoas tão especiais, tão longe de casa, que se tornaram uma família multiétnica, não tem preço! Sou muito feliz por conhecer pessoas tão competentes e de caráter tão engrandecedor. Obrigada por me ajudarem a moldar a pessoa que sou hoje.

Agradeço aos meus amados pais, José de Araújo Filho e Terezinha Damaceno Araújo, que nunca deixaram de acreditar em mim, nunca mediram esforços para me impulsionar nessa longa jornada e sempre se fizeram presentes, mesmo à distância. A eles devo tudo que sou.

RESUMO

A microscopia digital (MD) expandiu-se nos últimos anos em ambientes educacionais e profissionais para interconsulta, telepatologia, armazenamento e relatórios anatomopatológicos, colocando sistemas *whole slide imaging* (WSI) na posição privilegiada de dispositivos inovadores para interpretação de diagnósticos primários, aplicação previamente concebida com receio. Esta é uma consequência direta da falta de regulamentação desses dispositivos. É necessário reunir evidências sobre o desempenho da MD, para estabelecer se esta tecnologia pode ser usada para fornecer diagnóstico primário com segurança. O primeiro capítulo apresentado no presente estudo teve como objetivo fornecer informações sobre o desempenho de sistemas WSI, avaliando concordância intra-observador como melhor evidência. Uma busca eletrônica nas bases Scopus, MEDLINE/PubMed e Embase foi conduzida. As características metodológicas, a concordância entre a microscopia convencional (MC) e a MD e as razões para a ocorrência de diagnósticos discordantes foram analisadas. Um total de 13 artigos foram incluídos. As concordâncias intra-observadores variaram de 90% a 98,3% (intervalo de confiança de κ = 0,8-0,98). A dificuldade do caso foi o principal motivo de discordância (46,15%), seguido por dificuldades na identificação de microrganismos (15,38%). 58,84% enfatizam que o desempenho do método digital não está relacionado com a ocorrência de discordâncias. Apenas 25% das discordâncias tinham diagnósticos preferenciais por WSI. 15,38% dos estudos incluídos apresentaram alto risco de viés devido à seleção da amostra e 15,38% devido à ausência de especificação de um limiar de positividade. Todos os estudos foram classificados como baixa risco em relação à aplicabilidade. Esta revisão sistemática demonstrou uma alta concordância entre os diagnósticos por WSI e CLM. É possível confirmar que essa tecnologia pode ser usada para fornecer diagnóstico primário em várias especialidades da patologia humana. O segundo capítulo apresentado neste estudo teve como objetivo validar um sistema WSI para fins de diagnóstico de doenças bucais, utilizando a variabilidade intraobservador como a principal forma de análise. Setenta (n = 70) lâminas de vidro coradas em H&E de biópsias orais foram escaneadas pelo *Aperio Digital Pathology System (Aperio Technologies Inc., Vista, CA, EUA)* em uma magnificação de 20x. Dois patologistas experientes analisaram cegamente todos os casos com MLC e, após 3 meses de *washout*, com WSI. Informações clínicas foram fornecidas em ambas as análises. A concordância intraobservador entre os métodos foi de 97% para ambos os patologistas. Entre os casos discordantes, a maioria dos diagnósticos preferidos foi por MLC. Ambos os patologistas tiveram as mesmas discordâncias em diferentes casos. A dificuldade de alguns casos, que possibilitou interpretações controversas, e a pouca quantidade de tecido para análise foram consideradas razões principais de desacordo em detrimento dos métodos de diagnóstico. O valor de tempo (mediana) foi maior apenas com MLC para um patologista e, a melhoria do tempo com WSI está relacionada com o melhor fluxo de trabalho provido pelo sistema WSI. Os valores máximos de tempo ocorreram em casos discordantes e em outros casos considerados difíceis. Este estudo fornece evidências originais de um alto desempenho do sistema WSI para fins de diagnóstico na prática clínica, patologia de rotina e diagnóstico primário no campo da patologia oral.

Palavras-chave: Estudos de validação. Microscopia. Revisão sistemática. Boca-Doenças.

ABSTRACT

Digital microscopy (DM) has expanded recently in professional settings for interconsultations, telepathology, storage and routine reporting, what puts whole-slide imaging (WSI) systems in the privileged position of innovative devices for interpretation of primary diagnoses, application previously conceived with fear. This is a direct consequence of the lack of regulation of these devices. It is necessary to assemble evidence regarding the performance of the DM, in order to establish whether this technology can be used to provide primary diagnosis. The first chapter presented in this study aimed to provide information regarding the performance of whole slide imaging (WSI) devices, evaluating intraobserver agreement as the best evidence to elucidate whether digital microscopy (DM) is reliable for primary diagnostic purposes. Scopus, MEDLINE/PubMed and Embase were searched electronically. The methodological characteristics, the intraobserver agreement between conventional light microscopy (CLM) and WSI and the reasons for discordant diagnoses were analysed. Thirteen articles were included. The intraobserver agreements showed an excellent concordance, with values ranging from 90% to 98,3%, (κ coefficient range 0.8–0.98). Challenging cases were the main reasons for disagreements (46.15%) followed by difficulties in the identification of microorganisms (15.38%). 58,84% emphasize that the performance of the digital method is not related to the occurrence of disagreements. Only 25% of discordant cases had preferred WSI diagnosis. 15.38% presented high risk of bias due to unclear sample selection, and 15.38% due to the absence of specification of a threshold. Regarding to applicability, all studies were classified as a low concern. This systematic review showed a high concordance between diagnoses achieved by using WSI and CLM. These studies were also optimally designed to validate WSI for general clinical use and, most importantly, it is possible to confirm that this technology can be used to provide primary diagnosis in several specialties of human pathology.Second chapter of this study intended to validate a WSI system for diagnostic purposes of oral diseases, using the intraobserver variability as the primary form of analysis. Seventy (n = 70) H&E-stained glass slides of oral biopsies were scanned by the Aperio Digital Pathology System *(Aperio Technologies Inc., Vista, CA, USA)* at a magnification of 20x. Two experienced pathologists blindly analysed all cases with CLM and, after 3 months washout, with WSI. Clinical information was provided in both analyses. The intraobserver agreement between CLM and WSI system diagnoses was 97% for both pathologists. Among discordances, the majority of preferred diagnoses were by CLM. Both pathologists had the same discordances in different cases. Difficult cases, which allowed controversial interpretations, and the lack of tissue for analyses, were considered main reasons for disagreement rather than the diagnostic methods. Median time was higher only in CLM for one pathologist and the improvement of time in WSI was related to better workflow of WSI. Time outliers occurred in discordant cases and other difficult cases. This study provides original evidence for the high-performance of WSI for diagnostic purposes in clinical practice, routine pathology and primary diagnosis in the field of oral pathology.

Keywords: Validation studies. Microscopy. Review. Mouth - Diseases.

SUMÁRIO

1 INTRODUÇÃO

A microscopia digital (MD) está sendo difundida continuamente desde 1986. Esta tecnologia visa mimetizar o microscópio de luz convencional (MLC) por meio da utilização de imagens digitalizadas através de um emulador, que permite execução das mesmas funções possibilitadas por um MLC. Em 1999, foi desenvolvido, por Wetzel e Gilbertson, o primeiro sistema *whole slide imaging* (WSI) automatizado de alta resolução (Ho et al. 2006; Pantanowitz et al. 2011). Os sistemas WSI consistem em dois componentes, um software e um hardware, projetados para simular um MLC. Dessa forma é possível criar imagens digitais a partir do escaneamento de lâminas histológicas, citopatológicas e de imunohistoquímica e reproduzi-las na tela de um computador (Weinstein et al. 1987, 1989; Barker et al. 2001; Kayser et al. 2006; Yagi and Gilbertson 2007; Higgins 2015).

Devido à capacidade de substituir o MLC com extensa aplicabilidade, a MD fornece a possibilidade de renderizar diagnósticos mais precisos e representa um instrumento que encurta distâncias por meio do compartilhamento das imagens digitais para propósitos educacionais, de interconsultas e renderização de diagnóstico em localidades remotas (telepatologia). Além disso, seu uso compreende relatórios anatomopatológicos das lâminas coradas por meio da técnica da hematoxilina e eosina (H&E), interconsulta, interpretação e armazenamento das lâminas escaneadas, entre outros. Esta evolução, no entanto, também é responsável por distanciar o patologista da amostra real de tecido e, em alguns centros educacionais, responsável até mesmo pela extinção dos microscópios convencionais dos laboratórios de patologia. No entanto, isso não significa que o patologista será retirado de cena diante de um futuro totalmente digital (Weinstein et al. 2009; May 2010; Pantanowitz 2010; Park et al. 2012; Ghaznavi et al. 2013; Parwani et al. 2014; Boyce 2015; Fonseca et al. 2015).

Indubitavelmente, a patologia digital culmina em maior eficiência no fluxo de trabalho, maior acesso para serviços remotos, mais ergonomia e economia, já que vários departamentos podem custear um sistema WSI, economizando no custo de aquisição e, sobretudo, manutenção de múltiplos microscópios convencionais. No entanto, a adoção de um sistema WSI na rotina diagnóstica representa mais uma etapa a ser adicionada no processo (Dee 2009; Evans et al. 2009; Thorstenson 2009; Gabril and Yousef 2010; Hedvat 2010)

Apesar das vantagens, o custo do equipamento e a sua manutenção ainda são altos, embora os custos venham diminuindo com o passar do tempo. A qualidade da

imagem, a velocidade de aquisição da imagem, a manutenção dos arquivos digitais, os padrões e regulamentações referentes ao sistema WSI, bem como a possibilidade de uma performance inferior são motivos que afastam a MD da adesão universal, por parte dos profissionais, que precisam confiar totalmente em um sistema sem precedentes para renderizar diagnósticos primários em um serviço de rotina patológica (Gilbertson et al. 2006; Ho et al. 2006)

Além da preocupação principal sobre a confiabilidade do diagnóstico, existe ainda o receio em adotar um sistema que retarde o processo diagnóstico (Patterson et al. 2011). Em suma, os principais motivos para a relutância do uso desta tecnologia resumem-se aos altos custos de equipamento, ao tempo de escaneamento e à velocidade do microscópio virtual (Wienert et al. 2009).

Diante da expansiva utilização desta tecnologia e da falta de regulamentação para uso dos diferentes sistemas, em 2013, o *College of American Pathologists Pathology and Laboratory Quality Center* (CAP-PLQC), num esforço de delimitar recomendações que orientem os estudos de validação, elaboraram uma diretriz para a validação dos sistemas WSI (Pantanowitz et al. 2013). Em 2014, a *Canadian Association of Pathologists* elaborou diretrizes para estabelecer um serviço de telepatologia para patologia anatômica usando WSI (Bernard et al. 2014).

O *Food and Drug Administration* (FDA), responsável por regular os fabricantes de dispositivos eletrônicos, liberou o uso limitado de WSI para determinados tecidos, colorações e reagentes utilizados em imunohistoquímica (Cornish et al. 2012). Embora a FDA não tenha aprovado o uso dos sistemas WSI como substitutos do MLC para diagnósticos de rotina de patologia cirúrgica (Parwani et al. 2014), em abril de 2017, a FDA aprovou o primeiro sistema WSI que permite a revisão e interpretação de lâminas de patologia cirúrgica digital preparadas a partir de tecido biopsiado (*Food and Drug Administration*, 2017).

As informações provenientes de estudos de validação bem elaborados e baseados nas diretrizes disponíveis são muito importantes para guiar as agências que regulam os dispositivos WSI a proceder com a aprovação desses sistemas, quebrando as barreiras ainda existentes e comprovando a não inferioridade ou até mesmo a superioridade do método em detrimento do microscópio convencional (Bernard et al. 2014). Desta forma, espera-se que os argumentos que impedem a prática da patologia digital sejam rechaçados em prol deste grande avanço tecnológico para que, no futuro, a patologia totalmente digital torne-se uma realidade, figurando papel importantíssimo

para facilitar, agilizar e eliminar a variável interpretativa entre patologistas, assegurando um diagnóstico preciso e indubitável.

Este estudo foi elaborado com base nas diretrizes do CAP-PLQC e sugestões da DPA e propôs avaliar a variabilidade intra-observador entre o sistema CLM e WSI, como medida primordial para avaliar o desempenho do sistema WSI, para fins de diagnóstico de doenças orais na prática clínica, patologia de rotina e diagnóstico primário. Este estudo testou a hipótese de que o sistema WSI é um método confiável para diagnósticos de doenças bucais.

2 ARTIGOS

2.1 The performance of digital microscopy for primary diagnosis in human pathology: a systematic review.

Anna Luíza Damaceno Araújo[a], Lady Paola Aristizábal Arboleda[a], Natalia Rangel Palmier[a], Jéssica Montenegro Fonsêca[a], Mariana de Pauli Paglioni[a], Wagner Gomes-Silva[a], Cristhian Camilo Troconis[a], Alan Roger Santos-Silva[a].
[a] Oral Diagnosis Department, Semiology and Oral Pathology Areas, Piracicaba Dental School, University of Campinas (UNICAMP), Piracicaba, São Paulo, Brazil.

Corresponding Author
Alan Roger Santos-Silva DDS, MSc, PhD
Oral Diagnosis Department
Piracicaba Dental School, University of Campinas (UNICAMP)
Adress: Av. Limeira, 901, Bairro Areão, Piracicaba-SP, Brasil.
Postal code: 13414-903
e-mail: alan@unicamp.br
Telephone: +55 19 21065320

Orcid of the authors
Anna Luíza Damaceno Araújo - 0000-0002-3725-8051
Lady Paola Aristizábal Arboleda – 000-0002-6090-1893
Natália Rangel Palmier - 0000-0001-9588-6227
Jéssica Montenegro Fonsêca – 0000-0002-1217-7298
Mariana de Pauli Paglioni - 0000-0002-6588-9651
Wagner Gomes-Silva - 0000-0002-4219-0899
Christian Camilo Madrid Troconis – 0000-0003-4058-1447
Alan Roger Santos-Silva - 0000-0003-2040-6617

Ethical Responsibilities of Author Section:
All authors had substantial contributions to the conception, draft and design of this work, as well as participation of the acquisition, analysis and interpretation of data for the work. The final version of this work was approved for publication by all parts included. If there is a need, all author agrees to be accountable for any aspects of the work and we ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The authors also state that the material is original, has not been published elsewhere, and is being submitted only to the Virchows Archiv.

Manuscript word count: 4916

ABSTRACT

Validation studies of whole slide imaging (WSI) devices produces solid evidence regarding applicability of this technology. This study aimed to provide information regarding the performance of WSI devices, evaluating intraobserver agreement as the best evidence to elucidate whether digital microscopy (DM) is reliable for primary diagnostic purposes. Scopus, MEDLINE/PubMed and Embase were searched electronically. The methodological characteristics and percentages of agreement between conventional light microscopy (CLM) and WSI were also evaluated. In addition, this review proposed to elucidate the reasons for the occurrence of discordant diagnoses. A total of 13 articles were included. The intraobserver agreements showed an excellent concordance, with values ranging from 90% to 98,3%, (κ coefficient range 0.8–0.98). Challenging cases were the main reasons for disagreements (46.15%) followed by difficulties in the identification of microorganisms (15.38%). 58.84% dismissed the performance of the digital method (low magnification, image quality, technical limitations or failure of the method) as reasons for discordances. Preferred diagnoses were provided in 61.53% and, among these, only 25% had a majority of preferred WSI diagnosis. Concerning to quality assessment, 15.38% presented high risk of bias due to unclear sample selection, and 15.38% due to the absence of specification of a threshold. Regarding to applicability, all studies were classified as a low concern. In general, this systematic review showed a high concordance between diagnoses achieved by using WSI and CLM. These studies were also optimally designed to validate WSI for general clinical use and to provide primary diagnosis in several specialties of human pathology.

Keywords: Whole slide imaging, Intraobserver agreement, Systematic Review.

INTRODUCTION

Validation studies regarding the feasibility of whole slide imaging (WSI) systems have been conducted by pathology laboratories in a wide range of subspecialties to produce solid evidence and support the use of this technology for several applications, including primary diagnosis. The Guideline statement of College of American Pathologists Pathology and Laboratory Quality Center (CAP-PLQC) for WSI systems validation summarizes recommendations, suggestions and expert consensus opinion about the methodology of validation studies in an effort to standardize the process. This guideline encompasses the need to include a sample set of at least 60 cases for one application, and to establish a diagnostic concordance between digital and glass slides for the same observer (intraobserver variability) with a minimum washout period of 2 weeks between views [1]. Surprisingly, the recommendations do not suggest a consecutive or random selection of the cases and the need to blind the evaluators, but highlights that the viewing can be random or non-random.

Validation studies are, by definition, cross-sectional studies, and their designs have many methodological variations, which should be considered when evidences are assembled [2]. All these variations lead to skewed estimates about the test accuracy. The most important variation concerns to how the sample was selected, included and analyzed [3]. Some aspects regarding configuration, purpose of the test and the risks that prevent the test from serving your purposes may have been considered in validation studies, since performance may be influenced by analyses bias, reproducibility, washout period, response time, as well as size, scope and suitability of certain types of specimens. Besides that, learning curve and performance problems may be related to the method or to the pathologists [2]. Apparently, the order of analyses (digital or conventional), in this context does not affect the interpretation [3].

The most common bias in diagnostic studies is *verification bias*/detection bias/*work-up bias* (when the reference standard is not applied in all sample), *incorporation bias* (when index test and reference standard are not independent, what leads to overestimation of sensibility and specificity of the test), and *inspection bias* (when the tests is not blinded). The methodological characteristics should be individually evaluated by domain, which represents the way that the study was conducted [4].

The most common problems identified in the design of previous published validation studies are the cases selection (sample with a narrow range of subspecialty

specimens or a known malignant diagnosis) and the comparisons of the study results with a "gold standard"/consensus diagnosis/expert diagnosis instead of establishes the concordance by assessing the intraobserver agreement [5].

The FDA recently approved a WSI system for primary diagnosis purposes [6] and, despite the fact that this statement highlighted some assurance about the safety and feasibility of the digital system, only one device was tested and approved. Regardless this achievement, individual validation studies conducted by each laboratory, customized for each service and WSI system used, are still necessary and will provides the best evidence to attest the feasibility of digital pathology, especially if based on CAP-PLQC guidelines.

Given the absence of a broader collective agreement for the use of WSI in the human pathology context, it is necessary to assemble evidence regarding the performance of the digital microscopy, in order to establish whether this technology can be used to provide primary diagnosis. Therefore, this systematic review tested the diagnostic performances of WSI in human pathology. In addition, this review provided access to the main reasons for disagreement occurrences.

MATERIALS AND METHODS

The present systematic review was conducted following the Guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [7] and was registered with PROSPERO database under the protocol CRD42018085593. The review question defined was: "Is digital microscopy performance reliable for use in clinical practice and routine surgical pathology for diagnostic purposes as conventional microscopy?". The best evidence to answer this question is intraobserver agreement [1].

DEFINITION OF ELEGIBILITY CRITERIA

The eligibility criteria (Table 1) was elaborated based on 2 important recommendations and 1 suggestion established by CAP-PLQC guidelines [1]: the validation process should include a sample set of at least 60 cases for one application; the validation study should establish diagnostic concordance between digital and glass slides for the same observer (i.e., intraobserver variability); and a washout period of at least 2 weeks should occur between viewing digital and glass slides.

Table 1. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Validation cross-sectional study	Articles published in foreign language;
At least 60 cases*	Articles about telepathology, cytopathology or immunohistochemistry;
Intraobserver agreement	Sample with a known malignant diagnose;**
The concordance percentage or kappa index should be reported *	Articles with lack of information about how the sample was analyzed;
At least 2 weeks* washout period	Studies which the primary goal was not to examine diagnostic concordance between WSI and CLM;
	Studies which aimed to establish the intraobserver agreement but instead: used two different samples; in which each pathologist only performed diagnosis by one method; in which whole slide imaging diagnosis were compared to a consensus panel or original diagnosis (it is not intraobserver agreement)**

* CAP-PLQC Guidelines for WSI systems validation (Pantanowitz et al, 2013).
** (Cornish et al, 2012)

LITERATURE REVIEW

Recognizing the need to check if there are similar systematic reviews registered, executed, in progress or published with the same theme, the primary researcher (A.L.A.) conducted a previous literature review. A systematic review, in progress, with a similar proposal registered with PROSPERO in 2015 entitled “The diagnostic accuracy of digital microscopy: a systematic review”, under the protocol CRD42015017859, was identified. Two systematic reviews published were also found: “A systematic analysis of discordant diagnoses in digital pathology compared with light microscopy” [8] and “The Diagnostic Concordance of Whole Slide Imaging and Light Microscopy: A Systematic Review” [9]. Based on that findings, the research team decided to proceed with the present systematic review, since the methodology of the present review enhances the performance of well-designed studies supported by a solid guideline [1], which can provide much more reliable evidence about the utilization of WSI systems performance to provide primary diagnosis in human pathology than the previously published systematic reviews.

SEARCH STRATEGY

An electronic search was carried out on the databases: Scopus (Elsevier, Amsterdam, the Netherlands), MEDLINE (Medline Industries, Mundelein, Illinois) by PubMed platform (National Center for Biotechnology Information, US National Library of Medicine, Bethesda, Maryland) and Embase (Elsevier, Amsterdam, the Netherlands). Scopus was the first database used (for being an interdisciplinary basis and having article indexing intelligence) in order to align the keywords. The search strategy used was the following: [ALL (validation) AND ALL ("whole slide imag*)]. In sequence, the search was reproduced in the other databases. As result, 599 articles from Scopus, 132 from Embase and 115 from PubMed were retrieved. A hand searching was conducted in order to identify any eligible articles that may not have been retrieved by search strategy, but none was compatible with the eligibility criteria.

ARTICLE SCREENING AND ELIGIBILITY EVALUATION

Two reviewers (A.L.D.A. and A.R.S.S.) independently conducted the screening of articles by reading title and abstract and excluding articles that clearly do not fill the eligibility criteria. The assessment of eligibility was guided by a flow diagram drawn on phase two of the quality assessment. The two reviewers proceed with the reading in full text of the articles screened to identify the eligible articles, and all primary reasons for exclusions were registered for the composition of article selection flow. Rayyan QCRI was used as reference manager to perform the screening of the articles, exclusion of duplicates and registration of primary reason for exclusion [10].

EXTRACTION OF QUALITATIVE AND QUANTITATIVE DATA AND QUALITY ASSESSMENT

The data extraction was conducted by the primary researcher (A.L.A) and guided by a tailored extraction data form (Appendix I) a toll originally suggested by *The Cochrane Collaboration* [11]. The tailored tool has 5 sections: general information, eligibility, interventions participants and sample, methods, risk of bias assessment, applicability and outcomes. The section of 'risk of bias assessment' and 'applicability' was added based on the tailored QUADAS-2 (University of Bristol, Bristol, England), a tool designed to assess the quality of primary diagnostic accuracy studies. Specific guidance for each signaling question was produced and some signaling question, which does not apply to the review, were removed (Appendix II). Qualitative and quantitative

data were tabulated and processed in Microsoft Excel®. The studies identified in this review were highly heterogeneous in what concerns to WSI system utilized, magnification, number of pathologists involved, specimen type (subspecialty), washout time, and mainly how the sample was analyzed. These variations in studies design represents limitations and do not justify meta-analysis but only allow a narrative synthesis of the findings from the included studies.

RESULTS

PRISMA FLOWCHART

The search strategy identified a total of 846 records through database searching. After duplicates were removed, 681 records were screened and, among these, 48 articles were selected to be assessed for eligibility. A total of 13 articles [12–24] were included and 35 articles were excluded based on eligibility criteria. The composition of article selection flow is shown in Figure 1.

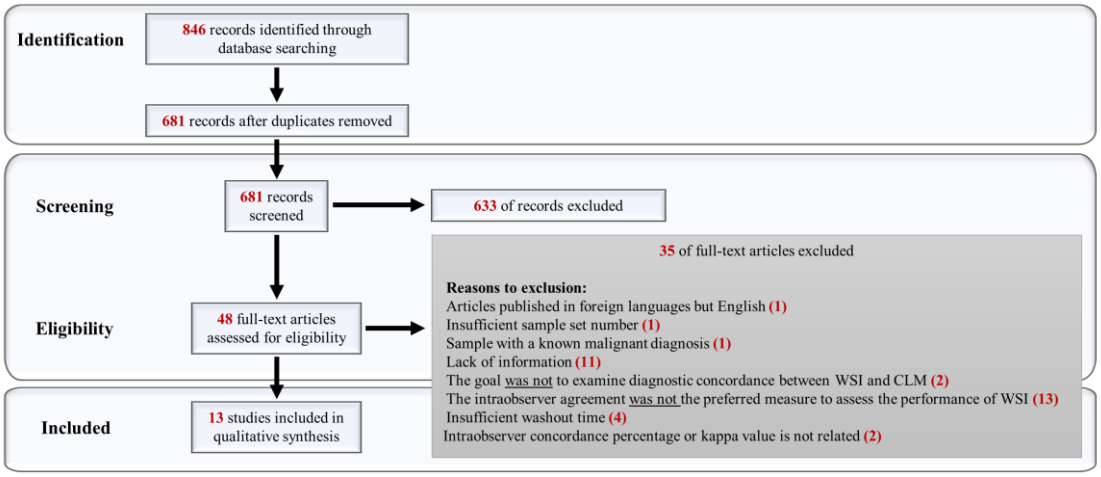


Figure 1. Flow Diagram of literature search adapted from PRISMA (Moher et al, 2009).

One article (2.08%) [25] was excluded for being published in French, 1 (2.08%) [26] for having insufficient sample set, 1 (2.08%) [27] for having a sample with a known malignant diagnosis and 11 studies (22.91%) [28–38] for presenting only abstracts (grey literature). Two studies (4.16%) [39, 40] were excluded because the main objective was not to examine diagnostic concordance between WSI and CLM. Four studies (8,33%) [41–44] were excluded because utilized insufficient washout time between the analyses.

The most important eligibility criteria pointed that the intraobserver agreement should be the preferred measure to assess the performance of digital

microscopy, according to CAP-PLQC guidelines [1]. Thirteen studies (27.08%) did not fit that criteria and were excluded for the following reasons: in 6 studies (12.5%) [45–50] the pathologists only assessed WSI and the concordance were reached by comparing WSI diagnosis with original glass slide diagnosis; in 4 studies (8.33%) [51–54] the WSI diagnosis was compared to a consensus panel diagnosis; in 1 study (2.08%) [55] two groups of students only assessed WSI and other only assessed glass slides; in 2 studies (4.16%) [56, 57] the sample analyzed was not the same in both methods.

Also, in 2 studies (4.16%) [58, 59] neither intraobserver concordance percentage or kappa value was reported. Disagreements were confronted and resolved by consensus.

METHODOLOGICAL CHARACTERISTICS OF THE STUDIES

Methodological characteristics of the studies are shown in Table 2.

Table 2. Methodological characteristics of the studies

No	Authors	Guideline	Aim of the study	WSI system specifications	Pathologists	Sample characteristics	Intraobserver agreement	Preferred diagnosis	Disagreements reason/Dismissed reason/Difficulties	Conclusion of the study
1	Al-Janabi et al (2012)a	-	To test the feasibility of using WSI for the diagnosis of skin specimens.	**Scanner:** ScanScope XT (Aperio Technologies Inc., Vista, CA, USA) **Magnification:** 20x **Monitor settings/resolution:** Samsung 245B (Samsung, Seoul, South Korea) displays of 24" (resolution of 19203×1200 pixels).	**Number:** 6 Experient	**Subspecialty/specimen type:** Dermatopathology (n = 100). **How sample was analysed:** The sample was diagnosed microscopically six months to one year previously (each pathologist assessed his own cases in WSI). The rediagnosis was done by the same pathologist who did the initial diagnosis. **Clinical information was provided?** Yes.	94% (95% CI = 0.87-0.97)	Disagreements: 6 WSI: 1 CLM : 5	**Reason for disagreement:** different interpretation of difficult or borderline cases. **Dismissed as a reason for disagreement:** magnification or WSI quality.	Primary histopathological diagnosis of skin biopsies and resections can be done digitally using WSI.
2	Al-Janabi et al (2012)b	-	To test the feasibility of whole slide images for diagnosis of gastrointestinal tract specimens.	**Scanner:** ScanScope XT (Aperio Technologies Inc., Vista, CA, USA) **Magnification:** 20x **Monitor settings/resolution:** not mentioned.	**Number:** 5 Experient	**Subspecialty/specimen type:** Gastrointestinal pathology (n = 100) **How the sample was analysed:** A complete set of well-focused WSIs that had been diagnosed light microscopically by 5 pathologists in 2009 were selected, to guarantee a washout period of 6 to 12 months. The same pathologists who did the initial diagnosis were asked to rediagnose their own cases on WSIs to exclude interobserver variation as much as possible. **Clinical information was provided?** Yes.	95% (95% CI = 0.89-0.98)	Disagreements: 5 WSI: 3 CLM: 2	**Reason for disagreement:** identification of microorganisms like Candida albicans, Helicobacter pylori, and Giardia lamblia was sometimes difficult. **Dismissed as a reason for disagreement:** higher magnification appears not to be very relevant (and it will need extra time and significantly more storage).	Histopathologic diagnosis of routine gastrointestinal biopsies and resections can be done well on WSIs acquired using today's scanning technology.
3	Al-Janabi et al (2012)c	-	To test the feasibility of digital slide image–based diagnosis of breast specimens.	**Scanner:** ScanScope XT (Aperio Technologies Inc., Vista, CA, USA) **Magnification:** not mentioned. **Monitor settings/resolution:** 24-in displays (Samsung, Seoul, South Korea) with 1920×1200 pixels.	**Number:** 1 Experient	**Subspecialty/specimen type:** Breast pathology (n = 100) **How the sample was analysed:** Specimens that had been diagnosed using light microscopy in 2008 to 2010 were selected to guarantee a washout period of at least 6 months. **Clinical information was provided?** Yes.	93% (95% CI = 86-97)	Disagreements: 4 WSI:4 CLM: 0	**Reason for disagreement:** borderline cases. **Dismissed as a reason for disagreement:** quality of digital slides.	This study demonstrates that upfront histopathologic diagnosis of breast biopsies and resections can reliably be done on digital slide image.
4	Al-Janabi et al (2013)	-	To evaluate the use of WSI for upfront diagnostics of placental tissue, and biopsies and resection from different body systems of patients under 18 years of age.	**Scanner:** ScanScope XT (Aperio Technologies Inc., Vista, CA, USA) **Magnification:** 20x **Monitor settings/resolution:** not mentioned.	**Number:** 1 Experient	**Subspecialty:** Paediatric pathology **Specimen type:** gastrointestinal, genitourinary, respiratory, skin, tonsil, grand and placentas. (n = 80) **How the sample was analysed:** These cases had been diagnosed by light microscopy by one pathologist in 2009. The same pathologist who did the original diagnosis was asked to rediagnose his own cases blinded to the original diagnosis on two other occasions; first digitally and then microscopically. The wash out time was more than 1 year. **Clinical information was provided?** Yes.	90% (95% CI = 0.84-0.96)	Disagreements: 10 WSI: 1 CLM: 9	**Reason for disagreement:** • Digital diagnosis of cases from the placenta was more time consuming (computer mouse is not the optimal tool for exploring WSI). • The identification of microorganisms like Candida albicans, Helicobacter pylori and Giardia lamblia was sometimes difficult. Missing microorganisms happened in one case. Scanning at 40× magnification would probably have given a more confident diagnosis of microorganisms. **Dismissed as a reason for disagreement**: higher magnification appears not to be very relevant (and it will need extra time and significantly more storage).	Histopathological diagnosis of biopsies and resections can generally be done well on WSI acquired using today's scanning technology. However, WSI scanned at 20× magnification was not optimal for exploring placental tissue.
5	Al-Janabi et al (2014)	-	To evaluate the feasibility of primary pathology diagnosis of urinary system specimens using WSI by comparing this to the performance when using a conventional microscopy.	**Scanner:** not mentioned **Magnification:** 20x **Monitor settings/resolution:** not mentioned.	**Number:** 2 Experient	**Subspecialty/specimen type:** Genitourinary pathology (n = 100) **How the sample was analysed:** WSI that had been conventionally diagnosed by two pathologists in 2008-2009 were selected. The same pathologists who did the initial diagnosis were asked to re-diagnose their own cases on WSI to exclude inter-observer variation as much as possible. Wash out ranged from 6 months to 1 year. **Clinical information was provided?** No.	87% (95% CI = 0.80-0.94)	Disagreements: 13 WSI: 6 CLM: 7	**Reasons for disagreement:** • WSI diagnosis task is more difficult and time consuming on 20x than on CLM; *No formal timing has been conducted. • Lack of clinical information; • Absent of feedback from multidisciplinary discussion; • Relative lack of routine, limited image resolution and suboptimal navigation tools;	Primary diagnostics of urinary tract specimens can be reliably done on WSI.
6	Arnold et al (2015)	CAP-PLQC	To determine the utility of CAP-PLQC guidelines to validating paediatric surgical pathology and cytopathology specimens.	**Scanner:** Aperio Model XT (Aperio Technologies Inc., Vista, CA, USA) **Magnification:** 20x or 40x **Monitor settings/resolution:** Dell monitors (Dell Corporation, Austin, TX, USA) with 1280-31024-pixel.	**Number:** 1 Previous training or experience was not mentioned	**Subspecialty**: Paediatric pathology **Specimen type:** liver, colon, oesophagus, stomach, placenta, skin, nerve, heart, colon, brain. (n = 473) - 60 surgical pathology cases, 130 specimen parts represented in 473 slides. **How the sample was analysed:** At the time of WSI review, all cases were at least 3 months from previous glass slide review. The resulting de-identified WSI cases were reviewed by the same paediatric pathologist who had previously completed clinical evaluation of the corresponding glass slides. **Clinical information was provided?** Yes.	98,3%	Disagreement: 1 WSI: 0 CLM:1	**Reasons for disagreement:** the difference in this diagnosis was primarily attributable to eosinophilic granular bodies, that were not detected by WSI review and identification of eosinophils and nucleated red blood cells varied between glass slide and WSI.	This study demonstrates that specimens representing the spectrum of paediatric surgical pathology practice can be reviewed using WSI.
7	Kent et al (2017)*	CAP-PLQC	To evaluate whether diagnosis from WSI on a digital microscope is inferior to diagnosis of glass slides from traditional microscopy (TM), with attention on image resolution, specifically eosinophils in inflammatory cases and mitotic figures in melanomas. To measure the workflow efficiency of WSI compared with TM.	**Scanner:** Aperio AT2 Image Scope (Aperio Technologies Inc., Vista, CA, USA) **Magnification:** 20x **Monitor settings/resolution:** not mentioned	**Number:** 3 Experient	**Subspecialty/specimen type:** Dermatopathology (n = 499) **How the sample was analysed:** Cases were divide in 3 groups. 3 board-certified dermatopathologists diagnoses one half of their cases by TM and the second by WSI. Glass slides were read on conventional microscopes, while WSI were read on an in-house WSI system. Following a minimum 30-day washout period, each dermatopathologist diagnosed the same cases using the alternative method. **Clinical information was provided?** Yes.	94%	-	**Reasons for disagreement:** the inherent subjectivity of pattern recognition and integration of degrees of dysplasia when biopsies are taken from chronically sun-damaged skin. **Dismissed as a reason for disagreement:** is not a failure of the WSI method.	Diagnosis from WSI was found to be noninferior compared with diagnosis from TM.
8	Loughrey et al (2015)*	CAP-PLQC	To evaluate primary digital pathology reporting in the setting of routine subspecialist gastrointestinal pathology. To compare individual digital and glass slide diagnoses.	**Scanner:** Hamamatsu Nanozoomer (Hamamatsu, United Kingdom) **Magnification**: 40x **Monitor settings/resolution:** not mentioned	**Number:** 3 Familiarized	**Subspecialty/specimen type:** Gastrointestinal pathology (n = 100) **How the sample was analysed:** The three study pathologists each independently evaluated by routine light microscopy all haematoxylin and eosin (H&E)-stained glass slides from each case accompanied by the patient demographic details and clinical information. After a washout period of at least 6 months, each of the three study pathologists independently evaluated the whole H&E-stained digital slide images for each of the 100 cases, with the same clinical information as provided for glass slide evaluation. **Clinical information was provided?** Yes.	95%	Disagreements: 14 WSI: 4 CLM:10	**Reasons for disagreement:** borderline calls (considered similarly likely to occur in digital or glass slide practice). **Dismissed as a reason for disagreement:** not related to anatomical segment within the gastrointestinal tract (oesophagus, stomach, duodenum, colorectum or appendix). **Difficulties which did not result in significant discordance:** • Image underexposure (due to scanning settings); • Examining WSI was perceived to take considerably more time than evaluation by conventional microscope. * No formal timing has been conducted.	The study provides further evidence to support validation of digital slide viewing as an alternative to light microscopy for primary reporting in the setting of gastrointestinal pathology.
9	Nielsen et al (2010)*	-	To investigate whether conventional microscopy of skin tumours can be replaced by virtual microscopy.	**Scanner:** Mirax Scan (Carl Zeiss MicroImaging, Göttingen, Germany) **Magnification:** 20x **Monitor resolution:** not mentioned	**Number:** 4 Trained	**Subspecialty/specimen type:** Dermatopathology (n = 96) **How the sample was analysed:** The digital slides were assessed first through intra hospital network connections using a virtual microscope that consisted of hard- and software supplied by Mirax. The digital slides were assessed twice with an intermediate time interval of at least 3 weeks. After at least 3 weeks, the conventional slides were assessed. This was done twice using traditional optical microscopes, again with an intermediate time interval of at least 3 weeks. **Clinical information was provided?** No.	κ = 0,93	-	**Reasons for disagreement:** individual interpretation, poor image quality, complexity of the cases and lack of clinical information.	It is feasible to make histologic diagnosis on the skin tumour types represented in this study using virtual microscopy.

Table 2. Methodological characteristics of the studies (continuation)

10	Pekmezci et al (2016)	CAP-PLQC	To assess the feasibility of primary pathology diagnosis of surgical neuropathology specimens using WSI.	**Scanner:** ScanScope XT (Aperio Technologies Inc., Vista, CA, USA) **Magnification:** 40x **Monitor resolution:** not mentioned	**Number:** 2 Experient	**Subspecialty/specimen type:** Neuropathology (n = 97) **How the sample was analysed:** The reviewers were expected to independently assess the virtual slides, render a diagnosis, and provide the WHO grade when applicable. Following a washout period of 2–6 months, both neuropathologists were provided with the original microscopic glass slides and the same clinical information used for WSI and the same parameters were recorded. **Clinical information was provided?** Yes.	Path 1: 94.9% Path 2: 88%		**Reasons for disagreement:** • Difficulties in the identification of mitotic figures in the WSI. • The loss of nuclear details and distortion of the chromatin pattern may at least partially explain some of the discordances. • In five cases, we were not able to identify any issue that may be associated with discrepancy (interpretive) nature than technical. • The need to narrow the diagnosis to one specific entity without being able to perform the above may be considered as a source for discordance in this study.	An all-encompassing conclusion about the utility of WSI for diagnostic purposes may not be available. We recommend independent validation for each subspecialty of pathology to identify subspecialty-specific concerns, so they can be properly addressed.
11	Saco et al (2017)*	-	To determine the accuracy of interpretation of WSI compared with conventional light microscopy in the diagnosis of needle liver biopsies.	**Scanner:** Ventana iScan HT (Ventana Medical Systems, Tucson, AZ, USA) **Magnification:** 400x **Monitor settings/resolution:** 30 Coronis fusion MDC4130 monitor 4 Megapixels (Barco Electronic Systems, Barcelona, Spain)	**Number:** 3 Experient	**Subspecialty/specimen type:** Liver pathology (n = 100) **How the sample was analysed:** Two experts analysed all cases. The first observer performed the initial evaluation with CLM, which was considered the reference for diagnostic attribution, and the second observation with WSI, whereas the second observer performed the initial evaluation with WSI and the second with CLM. An independent pathologist not involved with the evaluation compared the original CLM and the WSI-based evaluations and judged the concordance of the two diagnoses. **Clinical information was provided?** Yes.	Path 1: 96.6% k = 0.9 (95% CI: 0.9–1) Path 2: 90.3% k = 0.9 (95% CI: 0.8–0.9)	-	**Reasons for disagreement:** small size of the material or to intrinsic difficulty of the case. **Dismissed as a reason for disagreement:** none were related to **a poor quality of the WSI image or to insufficient magnification**	WSI can be safely used for primary histological diagnosis of liver biopsies, including native and transplantation specimens.
12	Tabata et al (2017)	CAP-PLQC	To demonstrate the availability of WSI-based primary diagnosis compared to light microscopy-based diagnosis.	**Scanners, magnifications and monitor resolutions:** IntelliSite Ultra Fast Scanner (Phillips Health, Amsterdam, Netherlands), 40x, 0.25mm/pixel; Aperio AT2 Scanner (Leica Biosystems, San Diego, CA, USA), 20x, 0.5 mm/pixel; NanoZoomer 2.0-HT C9600-13 (Hamamatsu photonics, Hamamatsu, Shizuoka, Japan), 20x, 0.46 mm/pixel; NanoZoomer 2.0-RS C10730-13 (Hamamatsu photonics), 20x, 0.46mm/pixel; NanoZoomer 2.0-RS C10730-13 (Hamamatsu photonics), 40 (0.23mm/pixel) VS800 (Olympus Corporation, Tokyo, Japan), 40x, 0.185mm/pixel; FINO (CLARO, Hirosaki, Aomori, Japan), 40x, 0.25mm/pixel.	**Number:** 10 Trained	**Subspecialty/specimen type:** upper gastrointestinal tract, lower gastrointestinal tract, female genital organ, genitourinary organ, breast and endocrine, head and neck, skin, haematopoietic organ, haepatobiliary-pancreatic organ, soft tissue and bone. (n = 100) **How the sample was analysed:** At each institute, all colleagues performed primary diagnoses by WSI of haematoxylin and eosin (HE)-stained slides, which were collected from 100 sequential cases including biopsy cases or surgical specimens containing <5 blocks and reviewed by light microscopy after a >2-week washout time. After the washout interval of over 2 weeks, the same observers reviewed conventional glass slides and diagnosed them by ordinary light microscopy. **Clinical information was provided?** Not mentioned.	96% (95% CI = 94.2–96.8)	Discrepant cases: WSI: 1 CLM: 8 Minor discrepancies: WSI: 17 CLM: 20	It is difficult to determine whether the discordance rate depends on disagreement between the WSI and microscopic findings, or intraobserver disagreement of pathological diagnosis. * The study avoided image degradation utilizing a display 3840x2160.	The results of this study demonstrated that WSI had good performance and usefulness for primary diagnosis,
13	Thrall et al (2015)	CAP-PLQC	To examine the results of a validation study performed using the draft version of the WSI clinical validation guideline recently released by the College of American Pathologists.	**Scanner:** iScan Coreo Au **Magnification:** 20x **Monitor resolution:** 1280 3 1084 pixels	**Number:** 57 Trained	**Subspecialty/specimen type:** Haematopathology, Neuropathology, medical kidney, and transplant biopsies; n = 2 sets of 100 cases to validate 10 scanners (1000 examinations); **How the sample was analysed:** In total, 2 sets of 100 cases were identified. The first set was used in all 3 phases to validate 2 scanners each time (6 total), and the second set was added in phases 2 and 3 to validate 2 scanners each time (4 total); The cases were given half as glass slides and half as images, with at least 3 weeks (21 days) before the cases were returned to be viewed again with the other modality. The glass slides given to the pathologists for review were the same representative slides that had been previously scanned for WSI. **Clinical information was provided?** Yes.	79%	-	**Reasons for disagreement:** • Insufficient attention to the critical foci; • Limited experience of the pathologists with WSI; • Uncarefully analysis by pathologists since these were not "real" cases with consequences to patients in the event of misdiagnosis; • WSI is disorienting and difficult to comprehensively analyse than glass slide under a microscope. • lack of image clarity at magnification above 320, which is an inherent limitation of the technology (becomes pixelated and unclear) • Increased concentration of challenging cases • Individual interpretation. **Dismissed as a reason for disagreement:** intraobserver variances do not derive from technical limitations of WSI.	The results were felt to validate the use of WSI for the intended applications in our multiinstitucional laboratory system.

***** Interobserver agreement were reported additional to intraobserver agreement

Included articles where published between 2010 and 2017. Six articles (46.15%) [14, 16, 21, 23, 24, 60] mentioned the use of CAP-PLQC guidelines, but the methodologies of all included studies were according to the established principles. The scanner manufacturer more commonly used was Scan Scope (Aperio, Vista, CA), which was reported in 8 studies (61.53%) [13, 14, 16–19, 21, 24].

The aims of the included studies were highly variable: 5 (38.46%) [13, 17, 18, 20, 24] aimed to test the feasibility, 2 (15.38%) [23, 24] aimed to determine the utility of CAP-PLQC guidelines [1], two (15.38%) [16, 60] intend to assess primary digital pathology reporting, 1 (7.69%) [15] proposed to determine the accuracy of WSI interpretation**,** 1 (7.69%) [12] proposed to investigate whether conventional microscopy of skin tumours can be replaced by virtual microscopy, 1 (7.69%) [14] proposed to evaluate whether diagnosis from WSI is inferior to diagnosis of glass slides and 1 (7.69%) [19] aimed to evaluate the use of WSI for diagnosis of placental tissue and paediatric biopsies.

Included studies performed validations in following areas: dermatopathology, hematopathology, neuropathology, gastrointestinal, genitourinary, breast, endocrine, soft tissue and bone, liver, head and neck and paediatric pathology areas. Transplant biopsies, hematopoietic and hepatobiliary-pancreatic organ biopsies are also included.

The median number of the samples was 100. The sample was analyzed in two different ways: (1) pathologists assessed digital slides or glass slides and, after a washout period, they reassessed the cases with the other modality; (2) when WSI diagnosis were compared to original glass slides diagnosis, the cases were address to the original pathologist, providing a satisfactory washout period and maintaining the intraobserver agreement as measure. One study (7.69%) [23] presented the first evaluation of half glass slides sample and half digital image sample with the analysis of the remaining samples by the opposite modality after washout. The washout period between views ranged from to 2 weeks to 12 months.

Three studies (23.07%) [12, 16, 23] reported set training and 8 (61.53%) reported previous experience of pathologists with WSI systems. One study (7.69%) [60] did not include a trained pathologist in the validation process but claimed that pathologist was familiar with the method. Previous training or experience was not mentioned in 1 study (7.69%) [21].

Only one 1 study (7.69%) [17] measured the scan time of slides (took on average 2.5 min) and only 1 (7.69%) [23] measured the diagnosis time (median time for glass slides was 132 seconds, and 210 seconds for WSI). Two studies (15.38%) [19, 20] considered WSI

more time consuming than CLM although no formal timing have been performed. A consensus diagnosis was mentioned to be used in 3 included studies (23.07%) [14, 16, 60].

INTRAOBSERVER CONCORDANCE

Among the included studies, 1 (7.69%) [12] did not report the percentage of concordance but related an almost perfect kappa index of 0,93. Two other studies (15.38%) [15, 24] reported their percentage of concordance for each pathologist, instead of overall concordance. For these reasons, these 3 studies were not graphically represented on the Figure 2. The majority of the intraobserver agreements reported showed an excellent concordance, with values ranging from 90% to 98,3%, (κ coefficient range 0.8–0.98). Only 1 study (7.69%) [23] showed a low concordance of 79%. All values of intraobserver agreement are shown in Table 2. Interobserver agreements were reported additionally to intraobserver agreement in 4 studies (30.76%) [12, 14, 15, 60].

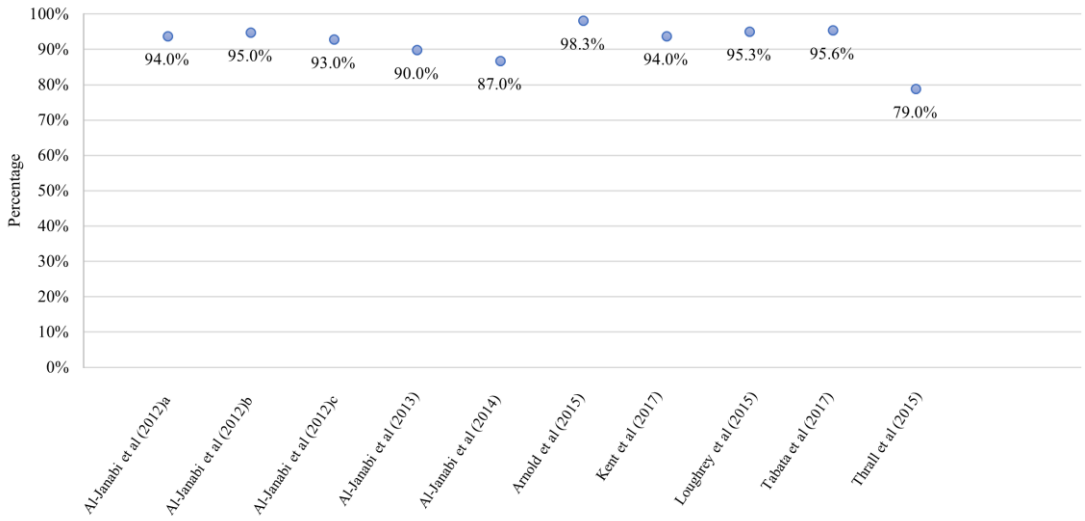


Figure 2. Graphic presentation of intraobserver agreement of included studies.

REASONS FOR DISAGREEMENTS

There were several reasons for disagreements, reported in conjunction, by each included study. It was possible to recognize reasons related to the case, to the pathologist, to the WSI system, to the way the test was conducted and to the glass slide quality. Among all related reasons that might explain the occurrence of discordant cases, the presence of borderline, difficult or challenging cases were the main reasons, reported in 6 articles (46.15%) [12, 15, 17, 18, 23, 60] followed by difficulties in the identification of microorganisms, related in 2 studies (15.38%) [13, 19]. One study (7.69%) [21] pointed

pitfalls in the identification of eosinophilic granular bodies, eosinophils and nucleated red blood cells, 1 (7.69%) [24] related difficulties in the identification of mitotic figures, nuclear details and chromatin patterns, 1 (7.69%) [14] pointed the inherent subjectivity of dysplasia, 2 (15.38%) [12, 20] reported lack of clinical information and 1 (7.69%) [15] indicated the small size of the material.

Other reasons for disagreements also reported were individual interpretation [12, 17, 23], non-optimal navigation tools [19, 20], the image resolution (poor image quality) and the lack of image clarity in higher magnifications (limitation of the technology, which becomes pixelated and unclear) [12, 20, 23], among others. Still, seven studies (58.84%) [13–15, 17–19, 23] dismissed the performance of the digital method (low magnification, image quality, technical limitations or failure of the method) as reasons for disagreement. Other disregarded reason was the anatomical segment, reported in 1 study (7.69%) [60]. The other 4 studies (30.79%) [12, 19, 20, 24] did not provide any information regarding this aspect and 1 (7.69%) [16] only limited to say it is difficult to determine whether the discordance depends on disagreement between the methods, or intraobserver disagreement of pathological diagnosis (it is possible the author intend to refers to the variations on the interpretations of pathological diagnosis, so intraobserver disagreement should not be used in this context).

Eight studies (61.53%) [13, 16–21, 60] provided the preferred diagnosis, when disagreements occurred. These diagnoses were reached reviewing the discordant cases and choosing the best diagnosis. Among those, only 2 studies (25%) [13, 18] had a majority of preferred WSI diagnosis and the need of a higher magnification and quality of digital slides were dismissed as reason for disagreement (the reasons were related to the difficulty in identifying microorganisms and borderline cases).

QUALITY ASSESSMENT (RISK OF BIAS)

The results of the quality assessment are shown in Table 3 and Figure 3. Among 13 included articles, 2 (15.38%) [13, 17] presented unclear risk of bias in sample selection due to unclear selection criteria. One study [24] excluded several lesions not relevant to the study (pituitary adenomas, degenerated diseases or other reactive lesions, metastatic carcinomas and melanomas, vascular malformations, and other benign or descriptive diagnoses such as meningocele, dermoid cyst, or focal cortical dysplasia) and also excluded cases for which the slides were not available for WSI scanning, which is acceptable and do not indicate bias. Two studies (15.38%) [23, 24] presented high risk of bias in the index test

due to the absence of specification of a threshold. The risk of bias was considered low in 100% of the other domains. Regarding to applicability, all studies were classified as a low concern.

Table 3. QUADAS-2 results

No.	Author	RISK OF BIAS				APPLICABILITY CONCERNS		
		Sample selection	Index test	Reference standard	Flow and timing	Sample selection	Index test	Reference standard
1	Al-Janabi et al (2012)a	?	☺	☺	☺	☺	☺	☺
2	Al-Janabi et al (2012)b	?	☺	☺	☺	☺	☺	☺
3	Al-Janabi et al (2012)c	☺	☺	☺	☺	☺	☺	☺
4	Al-Janabi et al (2013)	☺	☺	☺	☺	☺	☺	☺
5	Al-Janabi et al (2014)	☺	☺	☺	☺	☺	☺	☺
6	Arnold et al (2015)	☺	☺	☺	☺	☺	☺	☺
7	Kent et al (2017)	☺	☺	☺	☺	☺	☺	☺
8	Loughrey et al (2015)	☺	☺	☺	☺	☺	☺	☺
9	Nielsen et al (2010)	☺	☺	☺	☺	☺	☺	☺
10	Pekmezci et al (2016)	☺	☹	☺	☺	☺	☺	☺
11	Saco et al (2017)	☺	☺	☺	☺	☺	☺	☺
12	Tabata et al (2017)	☺	☺	☺	☺	☺	☺	☺
13	Thrall et al (2015)	☺	☹	☺	☺	☺	☺	☺

☺ Low risk
☹ High risk
? Unclear risk

Figure 3. Graphic presentation for QUADAS-2 results for included studies.

DISCUSSION

Validation studies have been improved among time and the recommendations of CAP-PLQC guidelines are particularly important on this aspect, since the standardization of the studies designs provides validations with homogeneous methodology [1]. The main purpose of systematic reviews is to minimize the chance of type I (systematic) error, by

eliminating studies with high risk of bias. Therefore, exclusion of highly discrepant methodologies studies allowed the comparison of only well-designed studies and the reaching of solid reliable conclusions. The way the sample is analyzed should encompass the index test and the reference standard with timing between analyses of paired samples (glass slide and correspondent digital slides). The analyses must be blinded, and the sample flow should encompass the analysis of all glass slides by CLM and, after the washout, the analysis of all digital slides.

Studies with a known malignant diagnosis (which may lead to a false high performance) and studies that compared WSI diagnosis with original or consensus diagnosis were excluded. These issues represents the most commons problems in validation studies [5] and generates selection bias [4]. The use of the index test only and the comparison with a consensus panel refers to a concept of accuracy, which is not a very recommended design for this particular purpose. Three articles included in this systematic review mentioned a consensus diagnosis in two different (and justifiable) situations: to include in the sample only appropriated cases to the intend purpose [14] and to reach a preferred diagnosis in discordant cases [16, 60]. The importance to reach a preferred diagnosis lies on the possibility to identify the pitfalls and missing details of the pathology, which are determinants in some cases [1].

Among included studies, 1 (7.69%) [15] proposed to determine the accuracy of WSI interpretation but presented intraobserver agreement instead. The accuracy is defined as concordance between the result of the method tested and the diagnosis established by a consensus or “gold standard”, while intraobserver agreement is basically the percentage of concordance between diagnosis reached by two different pathologists when they assess two diagnostic modalities [1]. The outcome of this study is not aligned with the aim but was founded to provide appropriated data, which allow the correct interpretation of the results. Another study [12] proposed to evaluate if diagnosis can be replaced by virtual microscopy and, for this purpose, the accuracy, sensitivity, specificity and predictive positive/negative values were measured. The accuracy, in this context, was defined as the addition of the percentage level of concordance and minor discordance, which is not the best concept definition. The diagnostic performance was intending to be calculated by means of sensitivity and specificity. However, sensitivity and specificity are used to calculate the reliability of the method and indicates the consistency of the results as the test is repeated, not the performance of the test. Fortunately, this study also provided the percentage of concordance (intraobserver agreement) between WSI and CLM diagnosis. It is very important to correctly delineate the

study design according to purpose. These sources of inconsistency generate divergent measures, sometimes not adequately to the purpose, and provide conflicting and not reliable data.

Validated pathology areas included dermatopathology, haematopathology, neuropathology, gastrointestinal, genitourinary, breast, endocrine, soft tissue and bone, liver, head and neck and paediatric pathology areas. Transplant biopsies, haematopoietic and haepatobiliary-pancreatic organ biopsies are also included. However, haematopathology, endocrine and bone and soft-tissue pathology areas have not been fully studied [61]. However, Saco and colleagues considered in 2016 that the areas of hematopathology, endocrine pathology and soft tissue and bone had not been fully studied [61]. Tabata and colleagues, in 2017, included specimens of soft tissue specimens and bone pathology in the sample but it is impossible to know how representative these specimens were, and a more targeted and specific validation is recommended. Saco and colleagues had also pointed out the need for validations in the Head and Neck area as there was only one study in this area. Fortuitously, our research group was able to publish a validation in oral pathology [62], adding original evidence of a high performance of WSI in this unexplored area. This study was not added to this systematic review because it was published after the search.

The washout time is highly variable in the literature, and there is no consensus of what period is more appropriated to avoid recall bias, since either an inferior or an overextended washout may produce bias due to the sample flow. A small period of washout may cause memorization bias in the test and a long washout may allow diagnostic criteria to change over time [12]. Surprisingly, this systematic review found that the study with the lowest intraobserver agreement has been conducted with one of the shortest washout period (3 weeks) [23]. This study also stated that intraobserver variations do not derive from technical limitations of WSI.

The inclusion of trained pathologists encompasses one of the recommendations of CAP-PLQC and appears to provide better concordance rates and minor diagnosis time [1]. Most pathologists are convinced that diagnoses on WSI systems are more time consuming. However, the learning curve [42, 63] and the utilization of suboptimal tools for navigation [19, 20] are the explanation for this extended analyses time and may be also related to the lack of confidence and experience of the pathologist in the WSI manipulation [64]. Two included studies [19, 20] pointed suboptimal navigation tools as reasons for disagreement and only one [19] corelates this technical particularity with a dispendious analyses. However, the correct

analyses of this information indicate there is an increase of the analyses time, not an increase of diagnostic discordance.

The scan time represents an extra step in the diagnosis process and one of the chief barriers to digital pathology acceptance, even more than the time required to render diagnosis [44]. The file size depends on magnification of scanning [65] and may impact the scan time, which is highly variable and difficult to include as a part of the validation because it does not provide~~s~~ a reproducible practice, what explains the absent of timing in the majority of validation studies. This extra step should be considered, however, as a part of the involving process, not as a disadvantage of the method, and must be adopted in further validation studies.

Higher intraobserver agreement is related to the high quality of digital slides and a better workflow provided by WSI systems [66], which appears to be more easily to navigate, instead of handling glass slides [67]. Some studies perceived that digital microscopy provides best definition of histologic images and configures the best method for identification of microscopic structures [68]. The intraobserver agreement values of the included studies were high and supported the high performance of the digital method and even the study with a lower intraobserver agreement [23] dismissed the technical limitations of WSI as reason for disagreement. However, it's important to be able to recognize when an overestimation of the performance of the test occurs. Validation studies have incorporation bias, since index test and reference standard are not independent. Besides that, intraobserver variability also increases when comparing the same glass slide overtime. Interobserver variability can also be increased in difficulty cases. This fact supports the cross-analysis of intra and interobserver variability [23]. However, CAP-PLQC advocated that it is important, for validation purposes to have one pathologist reproducing the same diagnosis with both modalities (i.e. intraobserver agreement) and the main objective is to accomplish a higher concordance rate [1]. The interobserver agreement should be avoided to evaluate the performance of the test because introduces bias due to the individual diagnostic interpretations [69].

The secondary objective of this review was to identify the reported reasons for disagreement and reach the cause of the problem, also stated by CAP-PLQC as an important outcome [1]. In this systematic review, the majority of the studies reported borderline cases as reasons for discordance occurrences and dismissed low magnification, quality of the image, technical limitations or failure of the method as reasons for disagreement. The difficulty caused by borderline cases is inherent of the method utilized and can occur in CLM as well

[60]. The subjectivity of some specimens (as dysplasia) [14] correlates directly to the experience and to individual interpretation of the pathologists. Studies which pointed quality of the digital images as reasons for discordant cases occurrence [12, 20, 23] also presents some other reasons that may contributed equally or even more for discordances occurrence as: lack of clinical information, borderline cases, inexperienced pathologists, individual interpretations, among others. Sometimes, there is a need of higher magnifications to conquer a better resolution to visualize subtle details which could be presented in difficult cases [65].

The impairment in recognizing eosinophilic granular bodies, eosinophils, mitotic figures or nuclear details and chromatin pattern, as well as some microorganisms, such as *Candida albicans*, *Helicobacter pylori*, and *Giardia lamblia* points to a limitation of the scanner, occur more frequently in some specifics subspecialty pathology areas (hematopathology, neuropathology, and gastrointestinal pathology). These pitfalls bring highlights to the need of more advanced scanners, which certainly should be improved with the advent of the technology improvement. Here lies the need of regulamentation of these devices, which should be standardized and improved. It is important to emphasize that, although difficulties in the identification of microorganisms were pointed as a reason for disagreement, higher magnifications were not considered to be very relevant by the authors [13, 19].

The lack of clinical information represents absent of reproducibility [1], increases difficulty in the diagnosis process and may lead to a wrong diagnose. Two included studies [12, 20] did not provide clinical data for the analysis and reported that this could made more difficult to render the diagnosis and may add an element of error [12] and the provision of clinical data may decrease these errors [20]. Other included study [16] did no mentioned if clinical data was provided and did not correlate disagreements with that fact or other possible factors. Fortunately, the majority of validation studies recognize the need of correlate the clinical characteristics and the histopathological to provide a correct diagnose, either through glass slide or digital slide.

The selection and inclusion of the cases should, ideally, be consecutively or random. However, it is known that this selection strategy may not provide a very representative sample with the most relevant diagnosis or a broad range of oral sites and tissue sources. A stratified uniform sampling is a better way to select the cases, as it gives smaller error estimation and may be useful to do measurements and estimates using cases grouped into strata [70]. Unfortunately, none of the included studies followed this methodology.

Besides, two studies included in this systematic review [13, 17] did not made clear how the samples were retrieved. An inappropriate exclusion of cases may result in overoptimistic estimates of diagnostic accuracy [4]. One included study related exclusions [24], which was were found to be acceptable and coherent with the proposal of the study. The pre-specification of the test threshold is important so there is no bias in interpreting the results, which could lead to an overoptimistic estimate of the test performance [71]. Two included studies [23, 24] did not mention the threshold previously, but one [23] mentioned to keep the threshold deliberately low to maximize the identification of discordances.

In general, the studies included in this systematic review showed a high concordance between diagnoses achieved by using WSI and CLM. In addition, these studies were also optimally designed to validate WSI for general clinical use and, most importantly, it is possible to confirm that this technology can be used to provide primary diagnosis in several specialties of human pathology, such as dermatopathology, haematopathology, neuropathology, gastrointestinal, genitourinary, breast, endocrine, soft tissue and bone, liver, head and neck (as well as oral pathology) and paediatric pathology areas. Transplant biopsies, haematopoietic and haepatobiliary-pancreatic organ biopsies are also included. The reported difficulties related to specific findings of certain areas of pathology reinforce the need for validation studies in some areas not fully studied, such as haematopathology, endocrine and bone and soft-tissue pathology areas.

COMPLIANCE WITH ETHICAL STANDARDS

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

We declare that the authors have no financial relationship with any commercial associations, current and within the past five years, that might pose a potential, perceived or real conflict of interest. These include grants; patent licensing arrangements, consultancies, stock or other equity ownership, advisory board memberships, or payments for conducting or publicizing our study.

Contributions

All authors had substantial contributions to the conception, draft and design of this work, (Anna Luíza Damaceno Araújo, Natália Rangel Palmier, Cristhian Camilo Troconis and Alan Roger Santos-Silva), as well as participation of the acquisition (Lady Paola Aristizábal Arboleda, Natália Rangel Palmier, Jéssica Montenegro Fonsêca, Mariana de Pauli Paglioni and Wagner Gomes-Silva), analysis (Anna Luíza Damaceno Araújo and Lady Paola Aristizábal Arboleda) and interpretation (Anna Luíza Damaceno Araújo, Cristhian Camilo Troconis and Alan Roger Santos-Silva) of data for the work. The final version of this work was reviewed and approved for publication by all parts included. Authors Anna Luíza Damaceno Araújo and Alan Roger Santos-Silva takes full responsibility for the work as a whole, including the study design, access to data and the decision to submit and publish the manuscript.

REFERENCES

1. Pantanowitz L, Sinard JH, Henricks WH, et al (2013) 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 137:1710–1722 . doi: 10.5858/arpa.2013-0093-CP
2. Pantanowitz L, Evans A, Pfeifer J, et al (2011) Review of the current state of whole slide imaging in pathology. J Pathol Inform 2:36 . doi: 10.4103/2153-3539.83746
3. Koch LH, Lampros JN, Delong LK, et al (2009) Randomized comparison of virtual microscopy and traditional glass microscopy in diagnostic accuracy among dermatology and pathology residents. Hum Pathol 40:662–667 . doi: 10.1016/j.humpath.2008.10.009
4. Whiting P, Harbord R, Kleijnen J (2005) No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol 5:19 . doi: 10.1186/1471-2288-5-19
5. Cornish TC, Swapp RE, Kaplan KJ (2012) Whole-slide Imaging. Adv Anat Pathol 19:152–159 . doi: 10.1097/PAP.0b013e318253459e
6. Food and Drug Administration (2017) FDA allows marketing of first whole slide imaging system for digital pathology. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm552742.htm. Accessed 16 Mar 2017
7. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred Reporting Items for

Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6:e1000097 . doi: 10.1371/journal.pmed.1000097

8. Williams BJ, DaCosta P, Goacher E, Treanor D (2017) A Systematic Analysis of Discordant Diagnoses in Digital Pathology Compared With Light Microscopy. Arch Pathol Lab Med 141:1712–1718 . doi: 10.5858/arpa.2016-0494-OA
9. Goacher E, Randell R, Williams B, Treanor D (2017) The Diagnostic Concordance of Whole Slide Imaging and Light Microscopy: A Systematic Review. Arch Pathol Lab Med 141:151–161 . doi: 10.5858/arpa.2016-0025-RA
10. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan—a web and mobile app for systematic reviews. Syst Rev 5:210 . doi: 10.1186/s13643-016-0384-4
11. (2017) Cochrane Effective Practice and Organisation of Care (EPOC). In: Data Collect. form. EPOC Resour. Rev. authors. http://epoc.cochrane.org/epoc-specific-resources-review-authors.
12. Nielsen PS, Lindebjerg J, Rasmussen J, et al (2010) Virtual microscopy: an evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors. Hum Pathol 41:1770–1776 . doi: 10.1016/j.humpath.2010.05.015
13. Al-Janabi S, Huisman A, Vink A, et al (2012) Whole slide images for primary diagnostics of gastrointestinal tract pathology: a feasibility study. Hum Pathol 43:702–707 . doi: 10.1016/j.humpath.2011.06.017
14. Kent MN, Olsen TG, Feeser TA, et al (2017) Diagnostic accuracy of virtual pathology vs traditional microscopy in a large dermatopathology study. JAMA Dermatology 153:1285–1291 . doi: 10.1001/jamadermatol.2017.3284
15. Saco A, Diaz A, Hernandez M, et al (2017) Validation of whole-slide imaging in the primary diagnosis of liver biopsies in a University Hospital. Dig Liver Dis 49:1240–1246 . doi: 10.1016/j.dld.2017.07.002
16. Tabata K, Mori I, Sasaki T, et al (2017) Whole-slide imaging at primary pathological diagnosis: Validation of whole-slide imaging-based primary pathological diagnosis at twelve Japanese academic institutes. Pathol Int 67:547–554 . doi: 10.1111/pin.12590
17. Al-Janabi S, Huisman A, Vink A, et al (2012) Whole slide images for primary diagnostics in dermatopathology: a feasibility study. J Clin Pathol 65:152–158 . doi: 10.1136/jclinpath-2011-200277
18. Al-Janabi S, Huisman A, Willems SM, Van Diest PJ (2012) Digital slide images for primary diagnostics in breast pathology: a feasibility study. Hum Pathol 43:2318–2325 .

doi: 10.1016/j.humpath.2012.03.027

19. Al-Janabi S, Huisman A, Nikkels PGJ, et al (2013) Whole slide images for primary diagnostics of paediatric pathology specimens: a feasibility study. J Clin Pathol 66:218–223 . doi: 10.1136/jclinpath-2012-201104
20. Al-Janabi S, Huisman A, Jonges GN, et al (2014) Whole slide images for primary diagnostics of urinary system pathology: a feasibility study. J Ren Inj Prev 3:91–6 . doi: 10.12861/jrip.2014.26
21. Arnold MA, Chenever E, Baker PB, et al (2015) The College of American Pathologists Guidelines for Whole Slide Imaging Validation are Feasible for Pediatric Pathology: A Pediatric Pathology Practice Experience. Pediatr Dev Pathol 18:109–116 . doi: 10.2350/14-07-1523-OA.1
22. Loughrey MB, Kelly PJ, Houghton OP, et al (2015) Digital slide viewing for primary reporting in gastrointestinal pathology: a validation study. Virchows Arch 467:137–144 . doi: 10.1007/s00428-015-1780-1
23. Thrall MJ, Wimmer JL, Schwartz MR (2015) Validation of multiple whole slide imaging scanners based on the guideline from the College of American Pathologists pathology and laboratory quality center. Arch Pathol Lab Med 139:656–664 . doi: 10.5858/arpa.2014-0073-OA
24. Pekmezci M, Uysal SP, Orhan Y, et al (2016) Pitfalls in the use of whole slide imaging for the diagnosis of central nervous system tumors: A pilot study in surgical neuropathology. J Pathol Inform 7:25 . doi: 10.4103/2153-3539.181769
25. Camparo P, Ramirez A, Claude V, et al (2009) Whole slide imaging in daily routine examination in a pathologic department: Experience of a military hospital network in Paris. Rev Fr Lab 38:49–55 . doi: RFL-01-2008-38-408-1773-035x-101019-200812623
26. Wang M, Liu S, Xie C, et al (2015) Making Primary Diagnosis on Liver Allograft Biopsies With Whole Slide Images - A Validation Study. Am J Clin Pathol 144:A168 . doi: https://doi.org/10.1093/ajcp/144.suppl2.168
27. Gage JC, Joste N, Ronnett BM, et al (2013) A comparison of cervical histopathology variability using whole slide digitized images versus glass slides: Experience with a statewide registry. Hum Pathol 44:2542–2548 . doi: 10.1016/j.humpath.2013.06.015
28. Bradshaw S, Driman D, Dupre M, et al (2013) Inter- and Intra-Observer Agreement in Diagnosing Dysplasia in Barrett's Esophagus: Comparison of Routine Glass Slide vs. Digital Image Examination. Lab Investig 93:471–489 . doi: 10.1038/labinvest.2013.36

29. Eccher A, Calio A, Colombari R, et al (2015) Validation of Digital Whole Slide Imaging According To the College of American Pathologists Guidelines in the Evaluation of Pre-Implant Kidney Biopsies. Lab Investig 95:499A . doi: 10.1038/labinvest.2015.25
30. Zeitouni J, Jorda M, Reyes C, Nadji M Validation of whole slide imaging for the first line diagnosis of prostate biopsies. Lab Invest 92:519A–520A
31. Gerhard R, Honorio A, Gentili A, et al Primary histopathological diagnosis using whole slide imaging (WSI): A validation study. Lab Invest 94:399A
32. Goodman S, Kandil D, Khan A Diagnosis of breast needle core biopsies using whole slide imaging. Lab Invest 94:399A
33. Hoffmann J, McGinnis L, Mafnas CT, et al (2016) Validation of Digital Whole Slide Imaging System for Intraoperative Breast Sentinel Lymph Node Touch Prep Analysis: A Single Institution Experience. Lab Investig 96:391–402 . doi: 10.1177/20101058110200S101
34. Maleeff BE (2014) Validation of a digital pathology whole slide imaging system. Microsc Microanal 20:1410–1411 . doi: 10.1017/S1431927614008782
35. Parimi V, Borys A, Zhou Y, et al Validation of whole frozen section slide image diagnosis in surgical pathology. Lab Invest 96:399A–400A
36. Sturm B, Fleskens S, Bot F, et al (2013) Larynx virtual microscopy validation study. Virchows Arch 463:109 . doi: 10.1007/s00428-013-1444-y
37. Sturm B, Mooi W, Creytens D, et al (2017) Validation of diagnosing melanocytic lesions on whole slide images- does z-stack scanning improve diagnostic accuracy? Virchows Arch 471:S15
38. Wilson I, Treanor D, Williams B (2017) Belfast Pathology 2017. 10th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 20-23 June 2017. J Pathol 243:S1–S41 . doi: 10.1002/path.4984
39. Lee JJ, Jedrych J, Pantanowitz L (2017) Validation of Digital Pathology for Primary Histopathological Diagnosis of Routine , Inflammatory Dermatopathology Cases. 0:1–7
40. Randell R, Ruddle RA, Thomas RG, et al (2014) Diagnosis of major cancer resection specimens with virtual slides: Impact of a novel digital pathology workstation. Hum Pathol 45:2101–2106 . doi: 10.1016/j.humpath.2014.06.017
41. Jara-Lazaro AR, Tan PH (2012) Comparing digital and optical microscopy diagnoses of

breast and prostate core biopsies. Pathology 44:46–48 . doi: 10.1097/PAT.0b013e32834e4254

42. Krishnamurthy S, Mathews K, McClure S, et al (2013) Multi-institutional comparison of Whole slide digital imaging and optical microscopy for interpretation of hematoxylin-eosin-stained breast tissue sections. Arch Pathol Lab Med 137:1733–39 . doi: 10.5858/arpa.2012-0437-OA
43. Rodriguez-Urrego PA, Cronin AM, Al-Ahmadie HA, et al (2011) Interobserver and intraobserver reproducibility in digital and routine microscopic assessment of prostate needle biopsies. Hum Pathol 42:68–74 . doi: 10.1016/j.humpath.2010.07.001
44. Williams BJ, Hanby A, Millican-Slater R, et al (2018) Digital pathology for the primary diagnosis of breast histopathological specimens: an innovative validation and concordance study on digital pathology validation and training. Histopathology 72:662–671 . doi: 10.1111/his.13403
45. Brunelli M, Beccari S, Colombari R, et al (2014) iPathology cockpit diagnostic station: Validation according to College of American Pathologists Pathology and Laboratory Quality Center recommendation at the Hospital Trust and University of Verona. Diagn Pathol 9: . doi: 10.1186/1746-1596-9-S1-S12
46. Campbell W, Lele S, West W, et al (2012) Diagnoses Rendered by Whole Slide Imaging (WSI) Alone Are Accurate for Use in a General Surgical Pathology Practice. Lab Investig 92:494–509 . doi: 10.1038/labinvest.2012.23
47. Campbell WS, Lele SM, West WW, et al (2012) Concordance between whole-slide imaging and light microscopy for routine surgical pathology. Hum Pathol 43:1739–1744 . doi: 10.1016/j.humpath.2011.12.023
48. Campbell WS, Hinrichs SH, Lele SM, et al (2014) Whole slide imaging diagnostic concordance with light microscopy for breast needle biopsies. Hum Pathol 45:1713–1721 . doi: 10.1016/j.humpath.2014.04.007
49. Ordi J, Castillo P, Saco A, et al (2015) Validation of whole slide imaging in the primary diagnosis of gynaecological pathology in a University Hospital. J Clin Pathol 68:33–39 . doi: 10.1136/jclinpath-2014-202524
50. Snead DRJ, Tsang YW, Meskiri A, et al (2016) Validation of digital pathology imaging for primary histopathological diagnosis. Histopathology 68:1063–1072 . doi: 10.1111/his.12879
51. Shah KK, Lehman JS, Gibson LE, et al (2016) Validation of diagnostic accuracy with

whole-slide imaging compared with glass slide review in dermatopathology. J Am Acad Dermatol 75:1229–1237 . doi: 10.1016/j.jaad.2016.08.024

52. Mills AM, Gradecki SE, Horton BJ, et al (2017) Diagnostic Efficiency in Digital Pathology. 00:1–7
53. Elmore J, Longton G, Pepe M, et al (2017) A randomized study comparing digital imaging to traditional glass slide microscopy for breast biopsy and cancer diagnosis. J Pathol Inform 8:12 . doi: 10.4103/2153-3539.201920
54. Fónyad L, Krenács T, Nagy P, et al (2012) Validation of diagnostic accuracy using digital slides in routine histopathology. Diagn Pathol 7:35 . doi: 10.1186/1746-1596-7-35
55. Foad AFA (2017) Comparing the use of virtual and conventional light microscopy in practical sessions: Virtual reality in Tabuk University. J Taibah Univ Med Sci 12:183–186 . doi: 10.1016/j.jtumed.2016.10.015
56. Bauer TW, Schoenfield L, Slaw RJ, et al (2013) Validation of whole slide imaging for primary diagnosis in surgical pathology. Arch Pathol Lab Med 137:518–524 . doi: 10.5858/arpa.2011-0678-OA
57. Buck T, Dilorio R, Havrilla L, O'Neill D (2014) Validation of a whole slide imaging system for primary diagnosis in surgical pathology: A community hospital experience. J Pathol Inform 5:43 . doi: 10.4103/2153-3539.145731
58. Bauer TW, Slaw RJ (2014) Validating whole-slide imaging for consultation diagnoses in surgical pathology. Arch Pathol Lab Med 138:1459–1465 . doi: 10.5858/arpa.2013-0541-OA
59. Mukhopadhyay S, Feldman MD, Abels E, et al (2017) Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology. Am J Surg Pathol 42:1 . doi: 10.1097/PAS.0000000000000948
60. Loughrey MB, Kelly PJ, Houghton OP, et al (2015) Digital slide viewing for primary reporting in gastrointestinal pathology: a validation study. Virchows Arch 467:137–44 . doi: 10.1007/s00428-015-1780-1
61. Saco A, Ramírez J, Rakislova N, et al (2016) Validation of Whole-Slide Imaging for Histolopathogical Diagnosis: Current State. Pathobiology 83:89–98 . doi: 10.1159/000442823
62. Araújo ALD, Amaral-Silva GK, Fonseca FP, et al (2018) Validation of digital microscopy in the histopathological diagnoses of oral diseases. Virchows Arch. doi:

10.1007/s00428-018-2382-5

63. Randell R, Ruddle RA, Mello-Thoms C, et al (2013) Virtual reality microscope versus conventional microscope regarding time to diagnosis: An experimental study. Histopathology 62:351–58 . doi: 10.1111/j.1365-2559.2012.04323.x
64. Sanders DSA, Grabsch H, Harrison R, et al (2012) Comparing virtual with conventional microscopy for the consensus diagnosis of Barrett's neoplasia in the AspECT Barrett's chemoprevention trial pathology audit. Histopathology 61:795–800 . doi: 10.1111/j.1365-2559.2012.04288.x
65. Romero Lauro G, Cable W, Lesniak A, et al (2013) Digital pathology consultations - A new era in digital imaging, challenges and practical applications. J Digit Imaging 26:668–677 . doi: 10.1007/s10278-013-9572-0
66. Boyce BF (2015) Whole slide imaging: Uses and limitations for surgical pathology and teaching. Biotech Histochem 90:321–330 . doi: 10.3109/10520295.2015.1033463
67. Vodovnik A (2016) Diagnostic time in digital pathology: A comparative study on 400 cases. J Pathol Inform 7:4 . doi: 10.4103/2153-3539.175377
68. Fernandes C, Bonan R, Bonan P, et al (2018) Dental Students' Perceptions and Performance in Use of Conventional and Virtual Microscopy in Oral Pathology. J Dent Educ 82:883–890 . doi: 10.21815/JDE.018.084
69. Fallon MA, Wilbur DC, Prasad M (2010) 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 134:1020–1023 . doi: 10.1043/2009-0320-OA.1
70. Särndal C-E (2003) Stratified Sampling. In: Model Assisted Survey Sampling. Springer, pp 100–109
71. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH (2008) Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions. Clin Chem 54:729–737 . doi: 10.1373/clinchem.2007.096032

THE COCHRANE COLLABORATION®

DATA COLLECTION FORM

Review title or ID

Study ID

Report IDs of other reports of this study *(e.g. duplicate publications, follow-up studies)*

1 GENERAL INFORMATION

Date form completed	
Name of person extracting data	
Report title	
Reference details *(author, year)*	
DOI/PUI/another ID of the publication	
Publication type *(full report, abstract, letter)*	

2 ELIGIBILITY

Study Characteristics	**Review Inclusion Criteria**	**Yes/ No/ Unclear**	**Values**	**Location in text**
Type of study	Cross sectional		-	
Sample set	At least 60 cases			
Types of intervention	Reference standards - conventional light microscopy (CLM) Index test– whole slide imaging system (WSI)		-	
Washout time	> 2 weeks			
Types of outcome measures	Was the intraobserver agreement the preferred measurement? (each observer assessing all sample by both methods – digital and conventional – with an appropriated wash out period between the analysis).			
	The percentage of concordance was reported?			
	Kappa index was reported?			
Decision:				
Reason for exclusion	Articles published in foreign languages but English. Insufficient sample set number.			

Study Characteristics	Review Inclusion Criteria	Yes/ No/ Unclear	Values	Location in text
	Sample with a known malignant diagnosis Studies with lack of information (mainly about how the sample was analysed); Studies which the primary goal was not to examine diagnostic concordance between WSI and CLM; The intraobserver agreement of the methods is the preferred measure to assess the performance of digital microscopy. Intraobserver concordance percentage or kappa value is not mentioned.			
Notes:				

DO NOT PROCEED IF STUDY EXCLUDED FROM REVIEW

3 INTERVENTIONS, PARTICIPANTS AND SAMPLE

	Description	Location in text
WSI system utilized and magnification of scanner:		
Computer settings/ monitor resolution:		
Pathologist number:		
Sample set quantity (n):		

4 METHODS

	Descriptions as stated in report/paper	Location in text
Type of study:		
The study was based on some stated Guideline?		
Aim of study:		
Pathologists were previous trained?		
How sample was analysed?		
Was there any information available along with the cases?		
Scan time or diagnosis time were measured?		
Washout time:		
Notes:		

5 RISK OF BIAS ASSESSMENT

Domain		Location in text
1. Sample selection **Describe methods of sample**		

selection: **Was the sample selection consecutive or random?** *Yes/No/Unclear* **A known malignant sample was avoided?** *Yes/No/Unclear* **Did the study avoid inappropriate exclusions?** *Yes/No/Unclear* **Could the selection of patients have introduced bias?** *RISK: LOW/HIGH/UNCLEAR*		*Risk of bias* *High (if at least one was reached as 'no' or 'unclear')* *Low (all reached as "yes"*	
Notes:			
2. Index test **Describe the index test and how is conducted and interpreted:** **Were the index test results interpreted without knowledge of the results of the reference standard?** *Yes/No/Unclear* **If a threshold (classification of the agreement) was used, was it pre-specified?** *Yes/No/Unclear* **Could the conduct or interpretation of the index test have introduced bias?** *RISK: LOW/HIGH/UNCLEAR*		*Risk of bias* *High (if at least one was reached as 'no' or 'unclear')* *Low (all reached as "yes"*	
Notes:			
3. Reference Standard **Describe the reference standard and how it was conducted and interpreted:**			

Domain			Location in text
1. Sample selection **Describe methods of sample selection:** **Was the sample selection consecutive or random?** *Yes/No/Unclear* **A known malignant sample was avoided?** *Yes/No/Unclear* **Is the reference standard likely to correctly classify the target conditions (diagnosis)?** *Yes/No/Unclear* **Were the reference standard results interpreted without knowledge of the results of the index test?** Yes/No/Unclear **Could the reference standard, its conduct, or its interpretation have introduced bias?** 1. *RISK: LOW/HIGH/UNCLEAR*			
		Risk of bias *High (if at least one was reached as 'no' or 'unclear')* *Low (all reached as "yes"*	
		Risk of bias *High (if at least one was reached as 'no' or 'unclear')* *Low (all reached as "yes"*	
Notes:			
4. Flow and timing **Describe the time interval and any interventions between index test(s) and reference standard:** **Could the patient flow have introduced bias?** *RISK: LOW/HIGH/UNCLEAR*			
Notes:			

6 APPLICABILITY

Domain		Location in text
1. Sample selection **Describe included cases (specimen type, subspecialty, biopsy location)?** **Is there a concern that the included cases do not match the review question?** *CONCERN: LOW/HIGH/UNCLEAR*		
Notes:		
2. Index test **Is there concern that the index test, its conduct, or interpretation differ from the review question?** *CONCERN: LOW/HIGH/UNCLEAR*		
Notes:		
3. Reference Standard **Is there concern that the reference standard, its conduct, or interpretation does not match with the review question?** *CONCERN: LOW/HIGH/UNCLEAR*		
Notes:		

7 OUTCOMES

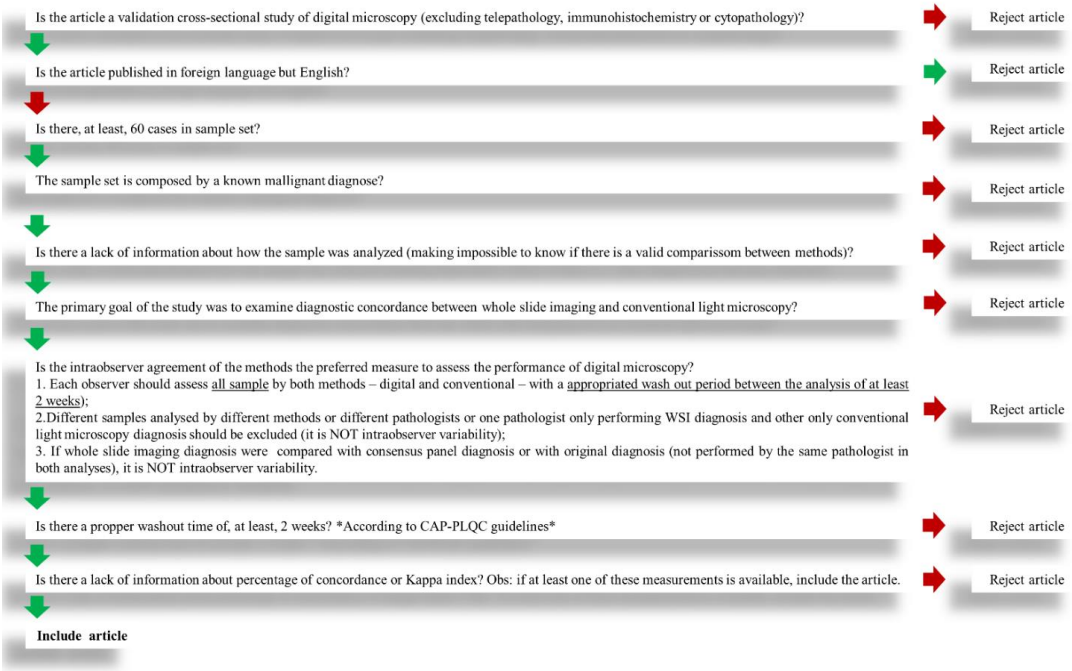
	Description as stated in report/paper	Location in text
Intraobserver agreement Concordance (%)		
Discordance (%)		
Kappa index		
Main reason for disagreement:		
In disagreement, what diagnosis were preferred?		
Discordant cases were reviewed?		
Conclusion of the study:		
Notes:		

APPENDIX II

TAILORED QUADAS-2

Phase 1. State the review question: The review question we intend to elucidate is: “Is digital microscopy performance reliable for use in clinical practice and routine surgical pathology for diagnostic purposes as conventional microscopy?”. For this purpose, we evaluate previous studies which compared digital microscopy (index test) with conventional microscopy (reference standard), in several pathology areas (target conditions) for diagnostic purposes (intended use). To assess the performance of whole slide imaging systems, we focus on intra-observer agreement (preferred measurement stated by CAP-PLQC guidelines). The sample set should include at least 60 cases and the pathologists involved on validation studies should perform an evaluation of all cases by two methods (conventional and digital) with a wash out period superior of 2 weeks. All these parameters obey the CAP-PLQC guidelines. Because the performance of the index test may depend on where it will be used in the diagnostic pathway, we should reinforce the need of a blind and independent analyses by both methods. Pathologists must assess either glass slides and correspondent whole slide images with a proper wash-out period (> 2 weeks). If WSI diagnosis were compared with original diagnosis (by glass slide), it is important that the digital slide be assessed by the same pathologist who made the original report (ensuring that the measure is intraobserver agreement).

Phase 2. Draw a flow diagram for the primary study:



Phase 3. Risk of bias and applicability judgments

Instructions:

Risk of bias (could be answer as: yes, no or unclear) - If all signaling questions for a domain are answered "yes" then the risk of bias can be judged "low." If any signaling issue is answered "no," this signals the potential for bias. The "unclear" category should only be used when insufficient data is reported to allow for judgment.

Applicability (could be answer as: low, high or unclear) - The applicability sections are structured similarly to the polarization sections, but do not include signaling issues. The

review authors should record the information on which the applicability judgment is made and then assess their concerns that the study does not match the review question. The "unclear" category should be used only when insufficient data are reported to allow judgment.

DOMAIN 1: PATIENT SELECTION
A. Risk of Bias
Describe methods of sample selection:

Was a consecutive or random sample enrolled? Yes/No/Unclear
A study should, ideally, include selected samples consecutively or randomly - otherwise, it has the potential to bias. If the sample include both (consecutively/randomly and non-consecutively/non-randomly), the risk of bias may be considered "low" if the percentage of non-consecutively/non-randomly cases was less than 10% of the total number of cases. If the selection of the samples was not clear, this signaling question must be rated as "unclear"

A known malignant sample was avoided? Yes/No/Unclear
A known malignant sample may lead to a super estimation of diagnostic accuracy (Cornish et al, 2012).

Did the study avoid inappropriate exclusions? Yes/No/Unclear
Inappropriate exclusion may result in overoptimistic estimates of diagnostic accuracy. If the study excluded > 10% the sample with or without specific motives, exclusions must be considered inadequate. This limit was determined pragmatically.

Could the selection of patients have introduced bias? RISK: LOW/HIGH/UNCLEAR

B. Concerns regarding applicability
Describe included cases (specimen type, subspecialty, biopsy location)?

Is there concern that the included cases do not match the review question?
CONCERN: LOW/HIGH/UNCLEAR

DOMAIN 2: INDEX TEST(S)
A. Risk of Bias
Describe the index test and how it was conducted and interpreted:

Were the index test results interpreted without knowledge of the results of the reference standard? Yes/No/Unclear
Interpretation of the results of the index tests can be influenced by the knowledge of the standard reference results (Whiting et al, 2004). The bias potential is related to the subjectivity of the test and the order of the test. Studies needs do clearly report blindness to answer this question with 'yes'.

If a threshold (classification of the agreement) was used, was it pre-specified? Yes/No/Unclear
For this question to be answered with 'yes', the study needs to mention which type of threshold was used and clearly indicate that it was specified prior to the start of the study. Selecting the test threshold to optimize sensitivity and/or specificity may lead to overoptimistic estimates of test performance, which is likely to be poorer in an independent sample of patients in whom the same threshold is used (Leeflang et al, 2008).

Could the conduct or interpretation of the index test have introduced bias? RISK: LOW/HIGH/UNCLEAR

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?

Variations in test technology, execution, or interpretation may affect estimates of its diagnostic accuracy. If index tests methods vary from those specified in the review question there may be concerns regarding applicability.

CONCERN: LOW/HIGH/UNCLEAR

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Describe the reference standard and how it was conducted and interpreted:

Is the reference standard likely to correctly classify the target condition (diagnosis)? Yes/No/Unclear

Estimates of test accuracy are based on the assumption that the reference standard is 100% sensitive and specific disagreements between the reference standard and index test are assumed to result from incorrect classification by the index test (Biesheuvel, Irwig and Bossuyt, 2007; van Rijkom and Verdonschot, 1995).

Were the reference standard results interpreted without knowledge of the results of the index test? Yes/No/Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias?

Potential for bias is related to the potential influence of prior knowledge on the interpretation of the reference standard (Whiting et al, 2004).

RISK: LOW/HIGH/UNCLEAR

B. Concerns regarding applicability

Is there concern that the reference standard, its conduct, or interpretation does not match with the review question?

CONCERN: LOW/HIGH/UNCLEAR

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Describe the time interval and any interventions between index test(s) and reference standard:

Could the sample flow have introduced bias? RISK: LOW /HIGH/UNCLEAR

References

1. *Biesheuvel C, Irwig L, Bossuyt P. Observed differences in diagnostic test accuracy between patient subgroups: is it real or due to reference standard misclassification? Clin Chem 2007; 53(10):1725-1729.*
2. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clinical Chemistry 2008; 54(4):729-737.
3. *van Rijkom HM, Verdonschot EH. Factors involved in validity measurements of diagnostic tests for approximal caries--a meta-analysis. Caries Research 1995; 29(5):364-70.*
4. *Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004; 140(3):189-202.*

2.2 Artigo: Validation of digital microscopy in the histopathological diagnoses of oral diseases

Artigo publicado no periódico Virchows Archiv (European Journal of Pathology)
DOI: 10.1007/s00428-018-2382-5 (Anexo 2)

Anna Luíza Damaceno Araújo[a], Gleyson Kleber do Amaral-Silva[a], Felipe Paiva Fonseca[b], Natália Rangel Palmier[a], Marcio Ajudarte Lopes[a], Paul M. Speight[c], Oslei Paes de Almeida[a], Pablo Agustin Vargas[a], Alan Roger Santos-Silva[a]

[a] Oral Diagnosis Department, Semiology and Oral Pathology Areas, Piracicaba Dental School, University of Campinas (UNICAMP), Piracicaba, São Paulo, Brazil.
[b] Department of Oral Surgery and Pathology, School of Dentistry, Federal University of Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil.
[c] Unit of Oral & Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, Sheffield, UK.

Corresponding Author
Alan Roger Santos-Silva DDS, MSc, PhD
Oral Diagnosis Department
Piracicaba Dental School, University of Campinas (UNICAMP)
Adress: Av. Limeira, 901, Bairro Areão, Piracicaba-SP, Brasil.
Postal code: 13414-903
e-mail: alan@unicamp.br
Telephone: +55 19 21065320

Orcid of the authors
Anna Luíza Damaceno Araújo - 0000-0002-3725-8051
Gleyson Kleber do Amaral-Silva - 0000-0002-6644-3264
Felipe Paiva Fonseca - 0000-0001-6989-0636
Marcio Ajudarte Lopes - 0000-0001-6677-0065
Paul Speight - 0000-0001-6459-1370
Pablo Agustin Vargas - 0000-0003-1840-4911
Alan Roger Santos-Silva - 0000-0003-2040-6617

Ethical Responsibilities of Author Section:
All authors had substantial contributions to the conception, draft and design of this work, as well as participation of the acquisition, analysis and interpretation of data for the work. The final version of this work was approved for publication by all parts included. If there is a need, all author agrees to be accountable for any aspects of the work and we ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The authors also state that the material is original, has not been published elsewhere, and is being submitted only to the Virchows Archiv.

Acknowledgements
The authors would like to gratefully acknowledge the financial support of the Coordination for the Improvement of Higher Education Personnel (CAPES/PROEX, Brazil), the National Council for Scientific and Technological Development (CNPq, Brazil) and the grants from São Paulo Research Foundation (FAPESP, Brazil) process number: 2009/53839-2, which supported the acquisition of the equipment used.

Manuscript word count: 2731

ABSTRACT

Whole slide imaging (WSI) systems are being increasingly used in educational and professional settings, highlighting the value of digital microscopy and favoring its acceptance for use in primary diagnosis. There has been a reluctance to introduce diagnostic applications due to a lack of validation and regulation of these devices. This study aims to provide information regarding the performance of WSI and to validate it for use in the diagnosis of oral diseases, using the intraobserver variability as the primary form of analysis. Seventy (n = 70) H&E-stained glass slides of oral biopsies were scanned using the Aperio Digital Pathology System at a magnification of 20x. Two experienced oral pathologists blindly analysed all H&E-stained sections with a conventional light microscope (CLM) and, after 3 months washout, with WSI. Clinical information was provided along with the cases in both analyses. The intraobserver agreement between CLM and WSI was 97% ($\kappa = 0.9$) for both pathologists. The majority of preferred diagnoses were by CLM. Both pathologists had the same discordances in different cases. Challenging cases, and cases with insufficient quantity of tissue for analyses were considered the main reasons for disagreement rather than the diagnostic methods. Median time taken to make a diagnosis was higher only in CLM for one pathologist. Time outliers occurred in discordant cases and in other difficult cases. This study provides evidence of a high-performance of WSI for diagnostic purposes in clinical practice, routine pathology and primary diagnosis in the field of oral pathology.

Keywords: Validation, Whole slide imaging, Digital pathology, Intraobserver agreement.

INTRODUCTION

Whole slide imaging (WSI) systems consist of devices to "convert" glass slides into multiple digital high-resolution images scanned by a camera. Software assembles all the images and enables them to be visualized as a single large image similar to a low power microscope view. It is also possible to magnify the image analogous to changing objective lenses [1]. The introduction of WSI is bringing about a paradigm shift in the way that we practice pathology. Over the last decade, WSI have been used for research, teaching, telepathology remote real-time interpretation of frozen sections and immunohistochemistry [2–4].

Major advantages of WSI are the possibility to analyse a slide from a remote access, share cases with experts and the inherent portability. In addition, WSI enables visualization of much more detail that the human eye is able to see by means of a conventional light microscope (CLM) [5]. WSI systems are more ergonomic, provide larger field of vision, easy navigation, allows a wider range of magnifications and make it possible to easily perform measurements and annotations [2]. These systems also provide high quality digital images, which enable conservation of cases and may prevent loss of data. Cloud storages eliminates storage problems, allow easy searching for case retrieval and put an end to the problems of broken glass slides and the inevitable fading of stains [1, 6].

However, there are still barriers to be overcome. The quality of the image, impediments to workflow, cost, threats to job security and the need for fast, high-capacity servers are some commonly cited disadvantages. Staining and focus may also be sensitive to the variations of the glass slide preparation. Badly positioned sections, chatter artefact, tissue folds and bubbles formed during coverslipping may result in poor focus and require a re-scan. A lack of familiarity with the technology increases time of diagnosis and hinders the workflow by slow performance [7–9]. Most studies have concluded that there is a learning curve, where the pathologists progressively improve their diagnosis time as they become familiarized with the technology [10, 11].

Due to the absence of recommendations to guide validation studies, the College of American Pathologists Pathology and Laboratory Quality Center (CAP-PLQC) have established guidelines for validation of WSI systems [12]. Subsequently, the Canadian Association of Pathologists released guidelines for establishing a telepathology service for anatomic pathology using WSI [13] and the Digital Pathology Association (DPA) has also provided additional criteria in this context [14]. The USA Food and Drug Administration

(FDA) is responsible for regulating device manufacturers and has approved limited use of WSI for some tissues, stains and reagents used in immunohistochemistry [7]. Recently, the FDA approved a WSI system via *de novo* classification, which is the only digital pathology system cleared for primary diagnostic use so far [15].

Given this scenario, it is necessary to provide validation of specific WSI systems before clinical use [16] and a re-validation when any significant change occurs [12]. Some groups are already using WSI in routine diagnostic services [5, 17, 18]. The most common problems in previous validation studies were the lack of research involving a large range of subspecialty specimens, comparisons of WSI diagnosis with a "gold standard" [7] and a sample containing known malignant diagnoses or challenging material [19, 20]. Regarding the current status of WSI systems validation, there have been studies on cytopathology, dermatopathology, neuropathology and gastrointestinal, breast, genitourinary, gynaecological, paediatric, pulmonary, renal, head and neck [2] and liver pathology areas [21]. However, there are still no studies published on oral pathology, hematopathology, endocrine, bone and soft-tissue pathologies. This lack of validation leads to a reluctance around the acceptance of the use of WSI [7].

Therefore, this study was designed based on the CAP-PLQC guidelines [12] and DPA suggestions [14] and proposes to evaluate intraobserver variability between CLM and WSI systems, as a measure to assess the performance of WSI systems, for diagnostic purposes of oral diseases in clinical practice, routine pathology and primary diagnosis. This study tested the hypothesis that WSI systems are a reliable method for diagnosis of oral diseases.

MATERIALS AND METHODS

STUDY DESIGN

This cross-sectional, retrospective study was approved by the Piracicaba Dental School/University of Campinas Ethics Committee in 05/06/2017 (registration: CAAE: 66762817.0.0000.5418). The sample consisted of seventy (n=70) H&E-stained glass slides of oral biopsies, randomly selected between the years 2002 and 2017, from a series of previously stipulated diagnoses, which aimed to cover the most common diseases in a routine oral pathology service, with a broad range of entities, oral sites and tissue sources. This approach aimed to avoid bias related to intrinsic diversity of cases and to improve variability, but also maintain equitability. The glass slides were scanned using the Aperio Digital Pathology System *(Aperio Technologies Inc., Vista, CA, USA)* with spatial sampling of 0.47 μm per

pixel, with automated focusing and magnification at 20x. All of the tissue present on glass slides were scanned and included in the digital images [12]. The monitor (Samsung, Seul, Coreia do Sul) used for slide viewing and interpretation had a screen resolution of 1600 x 900 pixels. Two pathologists, with extended previous experience in digital microscopy, blindly analysed and provided a diagnosis, in an independent way, for all cases with CLM, and after 3 months of washout, with WSI system. To achieve the recommendation of reproducibility [14], clinical information (age and sex of patients, anatomical site and clinical aspects of the lesions) was provided along with the cases. The diagnoses were compared between the two methods and classified as (1) concordant: diagnoses in both methods are the same; (2) slightly discordant: no clinical or prognostic implications or (3) discordant: with clinical/prognostic implications for the patient. Discordant cases were re-assessed to establish a preferred diagnosis between CLM and WSI in order to establish the reason for the disagreement, in particular to determine if discrepancies were due to factors in the method of preparation or to differences in the pathologists interpretation of the slides or images [21].

The pathologists involved descriptively pointed out technical problems in glass slides with the potential to affect the quality of the scanned images. The quality of glass slides and digital slides were stated as (1) poor: region of interest is compromised making diagnosis difficult or impossible; (2) diagnostic: insufficient tissue quantity, altered stain and/or deficiencies (artefacts or folds); (3) good: minor deficiencies (artifacts or folds) or (4) excellent: enough tissue quantity, appropriate stain, no artifacts or folds/whole material is focused, good color fidelity, no artefacts or folds [22]. Discordant cases were assessed in terms of quality to verify if this was an interfering factor for diagnostic concordance. The time taken to render a diagnosis was measured for each case, as an indicator of the workflow, since this factor is often used to resist the acceptance of digital methods [23].

STATISTICS

This study focused on the intraobserver agreement as the primary form of analysis and preferred measurement [12, 14]. We assessed Cohen κ statistics to establish the agreement between CLM and WSI (κ-values of < 0.00 were considered to indicate poor agreement, 0.0–0.2 slight agreement, 0.2–0.4 fair agreement, 0.4–0.6 moderate agreement, 0.6–0.8 substantial or good agreement, and >0.8 excellent or almost perfect agreement) [24]. The inter-observer variability was not explored. Statistical analyses were performed using VassarStats Website for Statistical Computation [25].

RESULTS

The oral diseases and oral sites are summarised in table 1.

Table 1. Included cases according to diagnoses and topography of the oral biopsies.

Range of lesions types and tissue sources	n (%)
Diagnoses	
Potentially malignant disorders	10 (14,28%)
Leukoplakia	10 (14,28%)
Actinic Cheilitis	
Epithelial malignant neoplasms	
Squamous cells carcinoma	10 (14,28%)
Minor/Major salivary glands, benign neoplasia	
Pleomorphic adenoma	10 (14,28%)
Minor salivary glands, malignant neoplasia	
Mucoepidermoid carcinoma	5 (7,14%)
Adenoid cystic carcinoma	5 (7,14%)
Odontogenic tumours	
Ameloblastoma, solid type	10 (14,28%)
Odontogenic cysts	
Odontogenic keratocyst	10 (14,28%)
Topography	
Floor of mouth	5 (7.14%)
Intraosseous	21 (30%)
Lower lip	10 (14.28%)
Upper lip	4 (5.73%)
Tongue	4 (5.73%)
Buccal mucosa	6 (8.57%)
Palate	12 (17. 14%)
Inferior alveolar ridge	5 (7.14%)
Superior alveolar ridge	2 (2.85%)
Retromolar trigone	1 (1.42%)
TOTAL	**70 (100%)**

Both pathologists had 68 concordant cases out of the 70 cases included in this validation study. The intraobserver agreement between CLM and WSI diagnoses was considered excellent ($\kappa = 0.967$; 95% CI: 0.876 – 1 for pathologist 1 and $\kappa = 0.967$; 95% CI: 0.877 – 1 for pathologist 2) with 97% agreement for both pathologists.

There were 2 discordant cases (with clinical/prognostic implications) for each observer, which were carefully analysed to elucidate the main reasons to disagreement. For pathologist 1, the WSI diagnosis was considered as correct in one case, whereas CLM diagnosis was judged as correct in the other. For pathologist 2, the CLM diagnosis was preferred in both cases (Table 2).

Table 2. Intraobserver discordant cases between methods, technical problems and correspondents preferred diagnosis.

Case no.	Pathologist 1		Pathologist 2		Technical problems	Preferred diagnosis
	CLM	WSI	CLM	WSI		
43	Actinic cheilitis	SCC	-	-	Insufficient quantity of tissue	SCC
55	-	-	ACC	Pleomorphic adenoma	Insufficient quantity of tissue	ACC
58	ACC	Pleomorphic adenoma	-	-	Insufficient quantity of tissue	ACC
65	-	-	SCC	Actinic cheilitis	-	SCC

SCC = squamous cell carcinoma; ACC = adenoid cystic carcinoma.

Technical problems used to measure the quality of the glass slides and the digital slides are presented in Table 3.

Table 3. Quality of glass slides and WSI.

	Glass slide		WSI	
	Pathologist 1	Pathologist 2	Pathologist 1	Pathologist 2
	n (%)	n (%)	n (%)	n (%)
Poor	-	-	-	-
Diagnostic	25 (35.71%)	14 (20%)	13 (18.57%)	17 (24.28%)
Good	-	-	3 (4.28%%)	-
Excellent	45 (64.28%)	56 (80%)	54 (77.14%)	53 (75.71%)

Glass slides criteria - Poor: region of interest is compromised making diagnosis difficult or impossible; diagnostic: insufficient tissue quantity, altered stain and/or deficiencies (artefacts or folds); good: minor deficiencies (artefacts or folds); excellent: enough tissue quantity, appropriate stain, no artefacts or folds.
WSI criteria - Poor: region of interests is compromised making diagnosis difficult or impossible; diagnostic: region of interests with blurred focus, altered stain and/or deficiencies (artefacts or folds); good: minor deficiencies (artefacts or folds); excellent: whole material is focused, good colour fidelity, no artefacts or folds.

Discordant cases were assessed in terms of quality to determine if this was an interfering factor for diagnostic concordance. Among four overall discordances, three presented insufficient quantity of tissue. Moreover, discordant cases involved the same diagnoses for both pathologists in different cases, and the spectrum of the cases allowed individual interpretations, which led to discordances. The discordances were also influenced by the complexity of cases.

Time to render a proper diagnosis was measured (Figure 1).

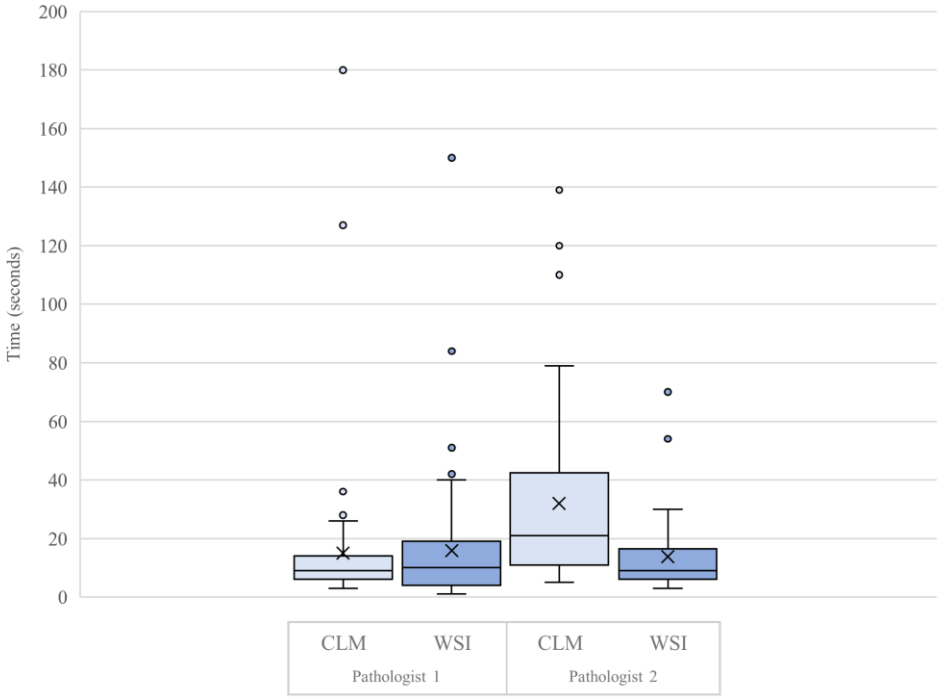


Figure 1. Box plot graphic with maximum, minimum, median and interquartile range of time needed for diagnoses for both pathologists in each method.

Similar median times were seen in both methods for pathologist 1 and in WSI for pathologist 2. Pathologist 2 showed a higher median time for CLM diagnoses, and an associated reduction of median time do render diagnoses by means of WSI. Among six cases with higher maximum time values for diagnoses, three were discordant cases (in a total of four overall discordances). The outlier time values occurred more frequently in cases of leukoplakia and adenoid cystic carcinoma (ACC) and were correlated to the inherent diagnostic difficulty of the cases (Table 4).

Table 4. Time to diagnosis outliers*.

Case no.	Diagnoses	Pathologist 1		Pathologist 2	
		CLM	WSI	CLM	WSI
05	Leukoplakia	481s*	104s*	90s	33s
30	Leukoplakia	6s	84s*	36s	11s
55**	ACC	127s*	150s*	110s*	56s*
58**	ACC	180s*	42s	120s*	21s
60	Leukoplakia	10s	33s	139s*	54s*
65**	~~Actinic cheilitis~~	12s	20s	120s*	14s

Discordant cases

DISCUSSION

This study represents the first validation of a WSI system used for histopathological diagnosis of oral diseases. The sample size (n = 70) is sufficient to cover spectrum and complexity of lesions usually observed in a routine practice, according to the recommendation of CAP-PLQC, which suggests that a sample set of at least 60 cases should be included in the validation process [12, 14]. Clinical information was provided along with the cases to reproduce the practice context [12, 14], as well as most of the well-designed published studies [21, 23, 26–29]. Additional H&E-stained slides, histochemical or immunohistochemical staining were not provided in any studied case to reach final diagnosis. Although clinical information has been provided, both pathologists pointed out that the absence of clinical photos and clinical diagnostic hypotheses represented limitations in the diagnostic process. The washout period chosen was of 3 months to minimize 'memorization bias'. This is a frequent variation in study design with most of the previously published studies stabilising a washout period between 2 weeks and 1 year [23, 26, 28, 30–34].

The best parameter to evaluate the performance of a WSI system against CLM is intraobserver agreement, rather than accuracy [12, 14, 34]. Intraobserver agreement refers to the percentage of diagnostic concordance when one observer assesses two methods with an interval of time although accuracy indicates the degree of agreement between the diagnosis result from the WSI and the "true diagnosis" (the one that is accepted, since it is established by a definition or consensus) [12] as a "gold standard". Some studies only compared the WSI with a gold standard [28], which represents a major problem in validation studies. The present study did not compare WSI with a gold standard.

Kappa statistics expressed the level of agreement between the methods and indicated an excellent concordance for both pathologists, similar to previously published studies [17, 21, 26, 31, 33, 35]. That may reflect the high quality of digital slides and better workflow of WSI [1]. Other studies were designed with different observers assessing each method (interobserver variability by instance) inserting an inevitable bias instead of only evaluating the performance of the method [36, 37]. The interobserver variability was not explored in this study since it is considered an expected variable due to the distinct interpretations of each pathologist and infer more about the pathologist experience and little about the method [36].

In discordant cases, the preferred diagnoses were agreed by review of CLM and WSI to verify which one provided the more coherent or "correct" diagnosis. In the present study, most of the preferred diagnoses (3/4) for discordant cases were obtained by CLM.

However, we recognize the need to analyse each case to assess if the discordances are related to the quality of WSI [6], intrinsic to the technology or due to other factors, since intraobserver variability can be increased even using the same glass slide over time [34].

Glass slide and correspondent digital slide quality were classified according to the presence of artefacts and folds, quantity of tissue, altered stain, blurred focus and good colour fidelity. Most cases were considered "excellent", and those classified as "diagnostic", were determined to provide enough material to render the diagnoses. Some studies did not consider the quality of digital images as a prominent cause of discordance[24], while others point it out as the main reason for diagnostic failure [6]. It is almost impossible to achieve optimum focus in entire digital image since tissue sections on a glass slide are very rarely planar [38]. In this study, three of the four discordant cases presented insufficient quantity of tissue for analyses.

Two cases (43 and 65) presented discordant diagnoses between actinic cheilitis and SCC, presenting areas of hyperkeratosis, acanthosis, atrophic epithelium, epithelial dysplasia, solar elastosis and microinvasion of epithelial cells in the lamina propria. In this context, the discordances may have occurred due to the fact that the pathologists did not observe the microinvasive areas or because these alterations may be interpreted as reactive epithelial atypia secondary to the lesion's inflammation rather than genuine dysplasia [39]. In these cases, the preferred diagnosis was judged as correct by CLM in one case and by WSI in the other, clearly disregarding the diagnostic method as the reason for the disagreement.

The other two discordant cases (55 and 58) involved discordances between pleomorphic adenoma and ACC, which are biphasic tumours that may present similar or overlapping morphological characteristics [40, 41]. These tumours often result in controversial interpretations and were considered difficult cases in the current study since both pathologists struggled to determine if the lesions were benign or malignant. The fact that both pathologists had the same discordances in different cases reinforces the possibility that these divergences are due to the difficulty of the cases, and variations of interpretations of each pathologist, rather than the diagnostic methods. In this study, the intrinsic difficulty of the cases influenced the occurrence of diagnostic discordances, rather than the method of preparation. In addition, there was a limited amount of tissue in these specimens, which is known to be a potential diagnostic pitfall in the differential diagnosis between these tumour types [42].

WSI offers a flexible viewing facility, requiring less time to identify histological structures and providing good definition [3], but the operation is influenced by the difficulty

and experience of handling and navigation, making time an important factor that reflects the workflow. In this study, the measurement of time to diagnosis was discrepant between pathologists. To allow a more coherent comparison, we assessed median values and concluded that median time was higher only in CLM for pathologist 2, not necessarily related to any difficult of the method. This result, when compared with WSI time for the same pathologist, indicates a reduction of time needed to render diagnoses using WSI, showing an improvement of the workflow. This may be related to better ergonomics, larger field of vision and full visualization as soon as the WSI was open, instead of glass slide handling [43]. This information disagrees with most published studies, which reported a range of 1 to 2 extra minutes of time required to render a diagnosis by virtual slides [4, 29, 44, 45]. Pathologist 1 presented a similar median time in both methods, also similar to the median time in WSI for pathologist 2.

For both pathologists the time outliers occurred more frequently in cases of leukoplakias and ACC. Discordant ACC cases presented insufficient quantity of tissue and the other outliers presented minimal technical problems (faded staining and tissue folding) insufficient to justify this range of exceeded time. The outlier's time values were higher for pathologist 1 with reduced time in two cases during WSI evaluation. Pathologist 2 presented a notable reduction of time to render the diagnoses in WSI.

In conclusion, this study provides original evidence of a high-performance for a WSI system in the histopathological diagnoses of oral diseases. Most importantly, the combination of a high concordance level between the studied methods and an outstanding workflow suggests that WSI is suitable for diagnostic purposes of oral diseases in clinical practice, routine pathology and primary diagnosis in the field of oral pathology. Therefore, this study accepted the hypothesis that a WSI system is a reliable method in oral diagnosis.

COMPLIANCE WITH ETHICAL STANDARDS

Funding:

This study was funded by the Coordination for the Improvement of Higher Education Personnel (CAPES/PROEX, Brazil), the National Council for Scientific and Technological Development (CNPq, Brazil) and the grants from São Paulo Research Foundation (FAPESP, Brazil) process number: 2009/53839-2, which supported the acquisition of the equipment used.

Conflict of interest

We declare that the authors have no financial relationship with any commercial associations, current and within the past five years, that might pose a potential, perceived or real conflict of interest. These include grants, patent licensing arrangements, consultancies, stock or other equity ownership, advisory board memberships, or payments for conducting or publicizing our study.

Contributions

All authors had substantial contributions to the conception (Anna Luíza Damaceno Araújo, Felipe Paiva Fonseca, Paul M. Speight and Alan Roger Santos-Silva), draft and design (Marcio Ajudarte Lopes, Oslei Paes de Almeida and Pablo Agustin Vargas) of this work, as well as participation of the acquisition (Natália Rangel Palmier and Gleyson Kleber Amaral-Silva), analysis (Oslei Paes de Almeida and Pablo Agustin Vargas) and interpretation (Anna Luíza Damaceno Araújo, Paul M. Speight and Alan Roger Santos-Silva) of data for the work. The final version of this work was reviewed and approved for publication by all parts included. Authors Anna Luíza Damaceno Araújo and Alan Roger Santos-Silva takes full responsibility for the work as a whole, including the study design, access to data and the decision to submit and publish the manuscript.

REFERENCES

1. Boyce BF (2015) Whole slide imaging: Uses and limitations for surgical pathology and teaching. Biotech Histochem 90:321–330 . doi: 10.3109/10520295.2015.1033463
2. Saco A, Ramírez J, Rakislova N, et al (2016) Validation of Whole-Slide Imaging for Histolopathogical Diagnosis: Current State. Pathobiology 83:89–98 . doi: 10.1159/000442823
3. Fonseca FP, Santos-Silva AR, Lopes MA, et al (2015) Transition from glass to digital slide microscopy in the teaching of oral pathology in a Brazilian dental school. Med Oral Patol Oral Cir Bucal 20:e17–e22 . doi: 10.4317/medoral.19863
4. Fine JL, Grzybicki DM, Silowash R, et al (2008) Evaluation of whole slide image immunohistochemistry interpretation in challenging prostate needle biopsies. Hum Pathol 39:564–572 . doi: 10.1016/j.humpath.2007.08.007
5. Parwani A, Pantanowitz L, Glassy E, Hassell L (2014) Regulatory barriers surrounding the use of whole slide imaging in the United States of America. J Pathol Inform 5:38 .

doi: 10.4103/2153-3539.143325

6. Stathonikos N, Veta M, Huisman A, van Diest P (2013) Going fully digital: Perspective of a Dutch academic pathology lab. J Pathol Inform 4:15 . doi: 10.4103/2153-3539.114206
7. Cornish TC, Swapp RE, Kaplan KJ (2012) Whole-slide imaging: Routine pathologic diagnosis. Adv Anat Pathol 19:152–159 . doi: 10.1097/PAP.0b013e318253459e
8. Pantanowitz L, Evans A, Pfeifer J, et al (2011) Review of the current state of whole slide imaging in pathology. J Pathol Inform 2:36 . doi: 10.4103/2153-3539.83746
9. Pantanowitz L, Farahani N, Parwani A (2015) Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. Pathol Lab Med Int 23 . doi: 10.2147/PLMI.S59826
10. Randell R, Ruddle RA, Mello-Thoms C, et al (2013) Virtual reality microscope versus conventional microscope regarding time to diagnosis: An experimental study. Histopathology 62:351–358 . doi: 10.1111/j.1365-2559.2012.04323.x
11. Krishnamurthy S, Mathews K, McClure S, et al (2013) Multi-institutional comparison of Whole slide digital imaging and optical microscopy for interpretation of hematoxylin-eosin-stained breast tissue sections. Arch Pathol Lab Med 137:1733–1739 . doi: 10.5858/arpa.2012-0437-OA
12. Pantanowitz L, Sinard JH, Henricks WH, et al (2013) 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 137:1710–1722 . doi: 10.5858/arpa.2013-0093-CP
13. Evans A, Garcia B, Godin C, et al (2014) Guidelines from the Canadian Association of Pathologists for establishing a telepathology service for anatomic pathology using whole-slide imaging. J Pathol Inform 5:15 . doi: 10.4103/2153-3539.129455
14. Digital Pathology Association Validating whole slide imaging for diagnostic purposes in pathology.http://www.cap.org/web/home/protocols-and-guidelines/cap-guidelines/current-cap-guidelines/validating-whole-slide-imaging-diagnostic-purposes?_afrLoop=72227620455902#!%40%40%3F_afrLoop%3D72227620455902%26_adf.ctrl-state%3D12evhanyqk_38. Accessed 16 Mar 2018
15. Food and Drug Administration (2017) FDA allows marketing of first whole slide imaging system for digital pathology.

    https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm552742.htm.

Accessed 16 Mar 2017

16. Lange H (2011) Digital Pathology: A Regulatory Overview. Lab Med 42:587–591 . doi: 10.1309/LMA2M9NQQF0ZCVHC
17. Evans AJ, Chetty R, Clarke BA, et al (2009) Primary frozen section diagnosis by robotic microscopy and virtual slide telepathology: the University Health Network experience. Semin Diagn Pathol 26:165–176 . doi: 10.1053/j.semdp.2009.09.006
18. Thorstenson S (2010) Digital pathology system. Case study. Advance Lab 19:69
19. Wilbur DC, Madi K, Colvin RB, et al (2009) Whole-slide imaging digital pathology as a platform for teleconsultation: a pilot study using paired subspecialist correlations. Arch Pathol Lab Med 133:1949–53 . doi: 10.1043/1543-2165-133.12.1949
20. Jukić DM, Drogowski LM, Martina J, Parwani A V (2011) Clinical examination and validation of primary diagnosis in anatomic pathology using whole slide digital images. Arch Pathol Lab Med 135:372–8 . doi: 10.1043/2009-0678-OA.1
21. Saco A, Diaz A, Hernandez M, et al (2017) Validation of whole-slide imaging in the primary diagnosis of liver biopsies in a University Hospital. Dig Liver Dis 49:1240–1246 . doi: 10.1016/j.dld.2017.07.002
22. Fónyad L, Krenács T, Nagy P, et al (2012) Validation of diagnostic accuracy using digital slides in routine histopathology. Diagn Pathol 7:35 . doi: 10.1186/1746-1596-7-35
23. Houghton JP, Ervine AJ, Kenny SL, et al (2014) Concordance between digital pathology and light microscopy in general surgical pathology: A pilot study of 100 cases. J Clin Pathol 67:1052–1055 . doi: 10.1136/jclinpath-2014-202491
24. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–74
25. Lowry R VassarStats:Web site for statistical computation. http://vassarstats.net/. Accessed 4 Mar 2018
26. Al-Janabi S, Huisman A, Vink A, et al (2012) Whole slide images for primary diagnostics in dermatopathology: A feasibility study. J Clin Pathol 65:152–158 . doi: 10.1136/jclinpath-2011-200277
27. Campbell WS, Hinrichs SH, Lele SM, et al (2014) Whole slide imaging diagnostic concordance with light microscopy for breast needle biopsies. Hum Pathol 45:1713–1721 . doi: 10.1016/j.humpath.2014.04.007
28. Bauer TW, Schoenfield L, Slaw RJ, et al (2013) Validation of whole slide imaging for primary diagnosis in surgical pathology. Arch Pathol Lab Med 137:518–524 . doi:

10.5858/arpa.2011-0678-OA

29. Van der Post RS, Van der Laak JAWM, Sturm B, et al (2013) The evaluation of colon biopsies using virtual microscopy is reliable. Histopathology 63:114–121 . doi: 10.1111/his.12131
30. Nielsen PS, Lindebjerg J, Rasmussen J, et al (2010) Virtual microscopy: An evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors. Hum Pathol 41:1770–1776 . doi: 10.1016/j.humpath.2010.05.015
31. Tabata K, Mori I, Sasaki T, et al (2017) Whole-slide imaging at primary pathological diagnosis: Validation of whole-slide imaging-based primary pathological diagnosis at twelve Japanese academic institutes. Pathol Int 67:547–554 . doi: 10.1111/pin.12590
32. Brunelli M, Beccari S, Colombari R, et al (2014) iPathology cockpit diagnostic station: Validation according to College of American Pathologists Pathology and Laboratory Quality Center recommendation at the Hospital Trust and University of Verona. Diagn Pathol 9:1–4 . doi: 10.1186/1746-1596-9-S1-S12
33. Frierson HF, Galgano MT (2007) Frozen-section diagnosis by wireless telepathology and ultra portable computer: use in pathology resident/faculty consultation. Hum Pathol 38:1330–1334 . doi: 10.1016/j.humpath.2007.02.006
34. Thrall MJ, Wimmer JL, Schwartz MR (2015) Validation of multiple whole slide imaging scanners based on the guideline from the College of American Pathologists pathology and laboratory quality center. Arch Pathol Lab Med 139:656–664 . doi: 10.5858/arpa.2014-0073-OA
35. Loughrey MB, Kelly PJ, Houghton OP, et al (2015) Digital slide viewing for primary reporting in gastrointestinal pathology: a validation study. Virchows Arch 467:137–144 . doi: 10.1007/s00428-015-1780-1
36. Fallon MA, Wilbur DC, Prasad M (2010) 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 134:1020–1023 . doi: 10.1043/2009-0320-OA.1
37. Ordi J, Castillo P, Saco A, et al (2015) Validation of whole slide imaging in the primary diagnosis of gynaecological pathology in a University Hospital. J Clin Pathol 68:33–39 . doi: 10.1136/jclinpath-2014-202524
38. Yagi Y GJ (2005) Speed, resolution, focus, and depth of field in whole slide imaging applications in clinical practice. In: Virtual Microscopy and Virtual Slides in Teaching,

Diagnosis, and Research Edited by: Gu J, Ogilvie RW. Boca Raton: Taylor & Francis;

39. Speight PM (2007) Update on oral epithelial dysplasia and progression to cancer. Head Neck Pathol 1:61–66 . doi: 10.1007/s12105-007-0014-5
40. Takeuchi J, Sobue M, Yoshida M, et al (1975) Pleomorphic adenoma of the salivary gland. With special reference to histochemical and electron microscopic studies and biochemical analysis of glycosaminoglycans in vivo and in vitro. Cancer 36:1771–89
41. Takeuchi J, Sobue M, Katoh Y, et al (1976) Morphologic and biologic characteristics of adenoid cystic carcinoma cells of the salivary gland. Cancer 38:2349–56
42. Khurram SA, Barrett AW, Speight PM (2017) Diagnostic difficulties in lesions of the minor salivary glands. Diag Histopathol 23:250–259 . doi: 10.1016/j.mpdhp.2017.04.008
43. Vodovnik A (2016) Diagnostic time in digital pathology: A comparative study on 400 cases. J Pathol Inform 7:4 . doi: 10.4103/2153-3539.175377
44. Weinstein RS, Descour MR, Liang C, et al (2004) An array microscope for ultrarapid virtual slide processing and telepathology. Design, fabrication, and validation study. Hum Pathol 35:1303–1314 . doi: 10.1016/j.humpath.2004.09.002
45. Li X, Liu J, Xu H, et al (2007) A feasibility study of virtual slides in surgical pathology in China. Hum Pathol 38:1842–1848 . doi: 10.1016/j.humpath.2007.04.019

3 DISCUSSÃO

A revisão sistemática (RS) acerca da performance dos dispositivos digitais para diagnósticos histopatológicos e o estudo de validação (EV) de WSI para diagnóstico histopatológico de doenças bucais apresentados neste trabalho seguiram as diretrizes do CAP-PLQC (Pantanowitz et al. 2013) e foram conduzidas concomitantemente. Desta forma, foi possível delinear o EV para que certos vieses fossem evitados, como por exemplo, o viés de verificação/detecção (que ocorre quando o padrão de referência não é aplicado em toda a amostra, apesar de o teste índice ter sido utilizado para toda a amostra) e o viés de inspeção (que ocorre quando o estudo não é cego). No entanto, o viés de incorporação é intrínseco à metodologia deste tipo de validação, que determina que as lâminas de vidro são correspondentes às lâminas digitais. Este pareamento gera dependência entre os resultados do teste (Whiting et al. 2005). Foi estabelecido que o presente EV obedeceria aos seguintes critérios: o número amostral seria superior a 60 e os casos abrangeriam as doenças bucais mais comuns e relevantes para um Serviço de Patologia Oral e Maxilofacial. Além disso, a amostra deveria ser analisada de modo que houvesse um intervalo de tempo apropriado (idealmente superior a duas semanas) entre a análise das lâminas de vidro (microscópio de luz convencional como teste padrão de referência) e das lâminas digitais (sistema WSI como teste índice). Esta análise deveria ser conduzida por patologistas treinados ou experientes, de forma cega e com fornecimento dos dados clínicos em ambas as análises (Pantanowitz et al. 2013).

Entende-se que essas recomendações visam padronizar as metodologias altamente variáveis dos estudos de validação, de modo que seja possível comparar estudos bem desenhados e os mais homogêneos possíveis, a fim de que seja possível fornecer evidências confiáveis. Estes mesmos critérios foram, então, utilizados como critérios de elegibilidade e guiaram a composição do fluxo de artigos incluídos na RS.

As áreas de patologia validadas incluíram áreas de Dermatopatologia, Hematopatologia, Neuropatologia, patologia gastrointestinal, geniturinária, de mama, de sistema endócrino, de tecidos moles e osso, de fígado, de cabeça e pescoço e patologia pediátrica. Biópsias de transplante, de órgãos hematopoiéticos e haepatobiliares-pancreáticos também estão incluídas. No entanto, as áreas de patologias hematológicas, endócrinas, ósseas e dos tecidos moles não foram integralmente investigadas do ponto de vista da validação (Saco et al, 2016).

* De acordo com as normas da UNICAMP/FOP, baseadas na padronização do International Committee of Medical Journal Editors - Vancouver Group. Abreviatura dos periódicos em conformidade com o PubMed.

O *washout* utilizado no EV foi de 3 meses. O tempo de *washout* é altamente variável em outros estudos disponíveis na literatura pertinente internacional, e não há consenso sobre qual período é mais apropriado para evitar viés de memória, uma vez que um *washout* menor ao usado ou muito maior pode produzir viés devido ao fluxo de análise da amostra. Um pequeno período de *washout* pode causar viés de memorização no teste e um *washout* longo pode permitir que critérios diagnósticos individuais mudem com o tempo (Nielsen et al. 2010).

O tempo das análises não é um componente da metodologia frequentemente reportado, embora necessário como um dos fatores para identificar parâmetros acerca do fluxo de trabalho de cada método. O EV identificou tempos similares entre os métodos para um patologista e redução de tempo, quando da análise digital, para o outro patologista, o que pode estar relacionado a uma melhor ergonomia, maior campo de visão e visualização completa e imediata, assim que o arquivo digital é aberto (Vodovnik 2016). No entanto, a curva de aprendizado (Krishnamurthy et al. 2013; Randell et al. 2013) e a utilização de ferramentas de navegação não otimizadas (Al-Janabi et al. 2013, 2014) explicam este longo tempo de análise e também podem estar relacionados à falta de confiança e experiência do patologista na manipulação do WSI (Sanders et al. 2012).

O melhor parâmetro para avaliar as performances dos sistemas digitais é a concordância intra-observador quando o mesmo observador analisa a amostra por meio de dois métodos diferentes (Pantanowitz et al. 2013). A concordância é expressa em porcentagem e pelo valor de κ com seus intervalos de confiança (Landis and Koch 1977). Esta medida nem sempre é utilizada ou relatada em EV, o que configurou um critério de inclusão importante para delimitar exatamente os estudos com as metodologias mais coesas, as quais a RS em questão se propôs a analisar. A concordância intra-observador quase perfeita entre MLC e WSI demonstrada no presente EV demonstrou alta qualidade das lâminas digitais e um melhor fluxo de trabalho acerca do método digital (Boyce 2015) que gera maior facilidade de navegação quando comparado ao manuseio das lâminas de vidro (Vodovnik 2016), oferece uma facilidade de visualização flexível (exigindo menos tempo para identificar estruturas histológicas) e fornece uma boa definição (Fonseca et al. 2015). No entanto, é importante reconhecer quando ocorre uma superestimação do desempenho do teste. Os estudos de validação têm viés de incorporação, uma vez que o teste de índice e o padrão de referência não são independentes. Além disso, a variabilidade intra-observador aumenta até mesmo

quando se compara a mesma lâmina de vidro *overtime*. A variabilidade inter-observador pode aumentar em casos difíceis. Este fato suporta a análise cruzada da variabilidade intra e inter-observador (Thrall et al. 2015). A variabilidade inter-observador foi pouco relatada dentre os estudos incluídos na RS e não foi explorada no EV desenvolvido como parte dessa dissertação, uma vez que é considerada uma variável que infere mais sobre a experiência do patologista e muito pouco sobre o método analisado (Fallon et al. 2010).

Todos os estudos incluídos na RS desenvolvida como parte dessa dissertação, bem como o EV propriamente dito, reconheceram a necessidade de analisar cada caso para avaliar se as discordâncias estavam relacionadas com a qualidade da lâminas digitais. O EV concluiu que as discordâncias diagnósticas ocorreram devido à dificuldade intrínseca dos casos e às variações de interpretação de cada patologista, em detrimento a dificuldades relacionadas aos métodos diagnósticos analisado. Adicionalmente, uma reduzida quantidade de tecido foi apontada como sendo uma potencial armadilha no diagnóstico diferencial entre tumores de glândula salivar menor (Speight 2007).

Da mesma forma, a maioria dos estudos incluídos na presente RS relatou casos limítrofes como razões para ocorrências de discordância, seguido de dificuldade técnica para permitir a visualização de microrganismos em cortes histológicos e rejeitou falha do método digital. É oportuno esclarecer que a dificuldade causada por casos limítrofes é inerente ao método utilizado e pode ocorrer também no MLC (Loughrey et al. 2015). É relevante mencionar que, apesar de dificuldades na identificação de microrganismos terem sido apontadas como motivo de discordância em estudos previamente publicados, a necessidade de equipamentos com conjuntos ópticos que permitem maiores magnificações não foi considerada relevante pelos autores (Al-Janabi et al. 2012, 2013).

* De acordo com as normas da UNICAMP/FOP, baseadas na padronização do International Committee of Medical Journal Editors - Vancouver Group. Abreviatura dos periódicos em conformidade com o PubMed.

4 CONCLUSÃO

Em geral, os estudos incluídos na RS que compõe o primeiro capítulo desta dissertação mostraram alta concordância entre os diagnósticos alcançados pelo uso de WSI e MLC. Esses estudos também foram idealmente projetados para validar WSI para uso clínico geral e é possível confirmar que esta tecnologia pode ser usada para fornecer diagnóstico primário em uma série de especialidades da patologia humana, como Dermatopatologia, Hematopatologia, Neuropatologia, além das áreas de patologia gastrointestinal, geniturinária, de mama, endócrina, de tecido moles e osso, de fígado, de cabeça e pescoço e pediátrica. Biópsias de transplante, de órgãos hematopoiéticos e hepatobiliares-pancreáticos também já foram incluídas em estudos de validação bem desenhados. As dificuldades descritas em relação aos achados específicos de determinadas áreas da Patologia reforçam a necessidade de estudos de validação específicos em áreas ainda não estudadas inteiramente, como Hematopatologia, patologia endócrina, óssea e de partes moles.

O segundo capítulo deste estudo fornece evidências originais de um alto desempenho do sistema WSI. A combinação de um alto nível de concordância entre os métodos estudados e um fluxo de trabalho otimizado sugere que o sistema WSI é adequado e seguro para fins de diagnóstico, sem desvantagens significativas que influenciem no diagnostico renderizado. Este estudo aceitou a hipótese de que o sistema WSI é um método confiável para diagnóstico histopatológico de doenças bucais.

* De acordo com as normas da UNICAMP/FOP, baseadas na padronização do International Committee of Medical Journal Editors - Vancouver Group. Abreviatura dos periódicos em conformidade com o PubMed.

REFERÊNCIAS

Al-Janabi S, Huisman A, Nikkels PGJ, ten Kate FJW, van Diest, PJ. Whole slide images for primary diagnostics of paediatric pathology specimens: A feasibility study. J Clin Pathol 2013;66(3):218–23 .

Al-Janabi S, Huisman A, Jonges GN, ten Kate FJW, Goldschmeding R, van Diest, PJ. Journal of Renal Injury Prevention Whole slide images for primary diagnostics of urinary system pathology: a feasibility study. J Ren Inj Prev J Ren Inj Prev 2014;3(34):91–96

Barker G, Krupinski EA, Larsen T, Erps K, Weinstein RS. Pay per view. The Arizona Telemedicine Program's billing results. TelemedicineJ E-Health 2001;7(4):287-91.

Bernard C, Chandrakanth SA, Cornell IS, Dalton J, Evans A, Garcia BM et al. Guidelines from the Canadian Association of Pathologists for establishing a telepathology service for anatomic pathology using whole-slide imaging. J Pathol Inform. 2014;5(1):15.

Boyce BF. Whole slide imaging: uses and limitations for surgical pathology and teaching. Biotech Histochem. 2015;90(5):321-30.

Cornish TC, Swapp RE, Kaplan KJ. Whole-slide imaging: Routine pathologic diagnosis. Adv Anat Pathol 2012;19(3):152-9.

Dee FR. Virtual microscopy in pathology education. Hum Pathol. 2009;40(8):1112-21.

Digital Pathology Association. [Accessed December 20, 2017]: Available from: http://www.cap.org/web/home/protocols-and-guidelines/cap-guidelines/current-cap-guidelines/validating-whole-slide-imaging-diagnostic-purposes?_afrLoop=72227620455902#!%40%40%3F_afrLoop%3D72227620455902%26_adf.ctrl-state%3D12evhanyqk_38

Evans AJ, Chetty R, Clarke BA, Croul S, Ghazarian DM, Kiehl TR, et al. Primary frozen section diagnosis by robotic microscopy and virtual slide telepathology: the University Health Network experience. Hum Pathol. 2009;40(8):1070-81.

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):1020–23

Fonseca FP, Santos-Silva AR, Lopes MA, Almeida OP, Vargas PA. Transition from glass to digital slide microscopy in the teaching of oral pathology in a Brazilian dental school. Med Oral Patol Oral Cir Bucal. 2015;20(1):e17.

* De acordo com as normas da UNICAMP/FOP, baseadas na padronização do International Committee of Medical Journal Editors - Vancouver Group. Abreviatura dos periódicos em conformidade com o PubMed.

Food and Drugs Administration. Washington: FDA [acesso 2018 03 jan]. Disponível em: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm552742.htm

Gabril MY, Yousef GM. Informatics for practicing anatomical pathologists: marking a new era in pathology practice. Mod Pathol. 2010;23(3):349-58.

Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: Whole-slide imaging and beyond. Annu Rev Pathol 2013;8(1):331-59.

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(1):4.

Hedvat CV. Digital microscopy: past, present, and future. Arch Pathol Lab Med. 2010;134(11):1666-70.

Higgins C. Applications and challenges of digital pathology and whole slide imaging. Biotech Histochem. 2015 Jul;90(5):341-7.

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):322–31

Kayser K, Molnar B, Weinstein RS. Digital pathology virtual slide technology in tissue-based diagnosis, research and education. Berlin:VSV Interdisciplinary Medical Publishing; 2006. p. 1-193.

Khurram SA, Barrett AW, Speight PM. Diagnostic difficulties in lesions of the minor salivary glands. Diagnostic Histopathol 2017:23:(6)250–59

Krishnamurthy S, Mathews K, McClure S, Murray M, Gilcrease M, Albarracin C et al. Multi-institutional comparison of Whole slide digital imaging and optical microscopy for interpretation of hematoxylin-eosin-stained breast tissue sections. Arch Pathol Lab Med 2013:137(12):1733-39.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977:33(1):159–74

Loughrey MB, Kelly PJ, Houghton OP, Coleman HG, Houghton JP, Carson A et al. Digital slide viewing for primary reporting in gastrointestinal pathology: a validation study. Virchows Arch 2015:467(2):137–44

May M. A better lens on disease. Sci Am 2010;10:74-7.

Nielsen PS, Lindebjerg J, Rasmussen J, Starklint H, Waldstrøm M, Nielsen B et al. Virtual microscopy: An evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors. Hum Pathol 2010:41(12):1770–76

Pantanowitz L. Digital images and the future of digital pathology. *Journal of Pathology Informatics*. 2010;1:15.

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.

Pantanowitz L, Sinard JH, Henricks WH, et al. College of American Pathologists Pathology and Laboratory Quality Center. 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 Dec;137(12):1710-22.

Park S, Pantanowitz L, Parwani AV. Digital imaging in pathology. Clin Lab Med 2012;32:557-84.

Parwani AV, Hassell L, Glassy E, Pantanowitz L. Regulatory barriers surrounding the use of whole slide imaging in the United States of America. *Jl of Pathol Inform* 2014;5:38.

Patterson ES, Rayo M, Gill C, Gurcan MN. Barriers and facilitators to adoption of soft copy interpretation from the user perspective: Lessons learned from filmless radiology for slideless pathology. J Pathol Inform. 2011 Jan 7;2:1.

Randell R, Ruddle RA, Mello-Thoms C, Thomas RG, Quirke P, Treanor D et al. Virtual reality microscope versus conventional microscope regarding time to diagnosis: An experimental study. Histopathology 2013:62:(2)351–58

Saco A, Ramírez J, Rakislova N, Mira A, Ordi J. Validation of Whole-Slide Imaging for Histolopathogical Diagnosis: Current State. Pathobiology. 2016;83(2-3):89-98.

Thrall MJ, Wimmer JL, Schwartz MR. Validation of multiple whole slide imaging scanners based on the guideline from the College of American Pathologists pathology and laboratory quality center. Arch Pathol Lab Med 2015:139(5):656–64.

Thorstenson S. From the conventional microscope to the digital slide scanner in routine diagnostic histopathology. Presented at Pathol. Vis. Conf., San Diego, 2009.

Vodovnik A. Diagnostic time in digital pathology: A comparative study on 400 cases. J Pathol Inform 2016;7(1):4

Weinstein RS, Bloom KJ, Rozek LS. Telepathology: system design and specifications. SPIE Proc Visual Common Image Process 1987; 845:404-7.

Weinstein RS, Bloom KJ, Rozek LS. Telepathology: long distance diagnosis. Am J Clin Pathol 1989;9:539-42.

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 Aug;40(8):1057-69.

Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol 2005:5(1):19

Wienert S, Beil M, Saeger K, Hufnagl P, Schrader T. Integration and acceleration of virtual microscopy as the key to successful implementation into the routine diagnostic process. Diagn Pathol. 2009 Jan 9;4:3.

Yagi Y, Gilbertson JR. Digital imaging in pathology: the case for standardization. J Telemed Telecare 2005;11:109-16.

COMITÊ DE ÉTICA EM PESQUISA
FACULDADE DE ODONTOLOGIA DE PIRACICABA
UNIVERSIDADE ESTADUAL DE CAMPINAS

CERTIFICADO

O Comitê de Ética em Pesquisa da FOP-UNICAMP certifica que o projeto de pesquisa "Validação da microscopia digital no diagnóstico histopatológico e citopatológico de doenças bucais", protocolo CAAE nº 66762817.0.0000.5418, dos pesquisadores Alan Roger dos Santos Silva, Pablo Agustin Vargas, Oslei Paes de Almeida, Felipe Paiva Fonseca e Anna Luíza Damaceno Araújo, satisfaz as exigências do Conselho Nacional de Saúde – Ministério da Saúde para as pesquisas em seres humanos e foi aprovado por este comitê em 05 de junho de 2017.

The Ethics Committee in Research of the Piracicaba Dental School, University of Campinas, certify that the project "Validation of digital microscopy in the histopathological and cytopathological diagnosis of oral diseases", CAAE nº 66762817.0.0000.5418, of Alan Roger dos Santos Silva, Pablo Agustin Vargas, Oslei Paes de Almeida, Felipe Paiva Fonseca and Anna Luíza Damaceno Araújo, comply with the recommendations of the National Health Council – Ministry of Health of Brazil for research in human subjects and therefore was approved by this committee on June 05, 2017.

Profa. Fernanda Miori Pascon
Vice Coordenador
CEP/FOP/UNICAMP

Prof. Jacks Jorge Junior
Coordenador
CEP/FOP/UNICAMP

Nota: O título do protocolo e a lista de autores aparece como fornecidos pelos pesquisadores, sem qualquer edição.
Notice: The title and the list of researchers of the project appears as provided by the authors, without editing.

Anexo 2 - Documento de aceite do artigo (PROOF)

JmlID 428_ArtID 2382_Proof# 1 - 25/05/2018

Virchows Archiv
https://doi.org/10.1007/s00428-018-2382-5

ORIGINAL ARTICLE

Validation of digital microscopy in the histopathological diagnoses of oral diseases

Anna Luíza Damaceno Araújo[1] · Gleyson Kleber Amaral-Silva[1] · Felipe Paiva Fonseca[2] · Natália Rangel Palmier[1] · Marcio Ajudarte Lopes[1] · Paul M. Speight[3] · Oslei Paes de Almeida[1] · Pablo Agustin Vargas[1] · Alan Roger Santos-Silva[1]

Received: 17 April 2018 / Accepted: 21 May 2018

Abstract
Whole slide imaging systems (WSIs) are being increasingly used in educational and professional settings, highlighting the value of digital microscopy and favouring its acceptance for use in primary diagnosis. There has been a reluctance to introduce diagnostic applications due to a lack of validation and regulation of these devices. This study aims to provide information regarding the performance of WSI and to validate it for use in the diagnosis of oral diseases, using the intra-observer variability as the primary form of analysis. Seventy ($n = 70$) H&E-stained glass slides of oral biopsies were scanned using the Aperio Digital Pathology System at a magnification of × 20. Two experienced oral pathologists blindly analysed all H&E-stained sections with a conventional light microscope (CLM) and, after 3-month washout, with WSI. Clinical information was provided along with the cases in both analyses. The intra-observer agreement between CLM and WSI was 97% ($\kappa = 0.9$) for both pathologists. The majority of preferred diagnoses were by CLM. Both pathologists had the same discordances in different cases. Challenging cases and cases with insufficient quantity of tissue for analyses were considered the main reasons for disagreement rather than the diagnostic methods. Median time taken to make a diagnosis was higher only in CLM for one pathologist. Time outliers occurred in discordant cases and in other difficult cases. This study provides evidence of a high performance of WSI for diagnostic purposes in clinical practice, routine pathology and primary diagnosis in the field of oral pathology.

Keywords Validation · Whole slide images · Digital pathology · Intra-observer agreement

Introduction

Whole slide imaging systems (WSIs) consist of devices to 'convert' glass slides into multiple digital high-resolution images scanned by a camera. Software assembles all the images and enables them to be visualised as a single large image similar to a low power microscope view. It is also possible to magnify the image analogous to changing objective lenses [1]. The introduction of WSI is bringing about a paradigm shift in the way that we practice pathology. Over the last decade, WSIs have been used for research, teaching, telepathology remote real-time interpretation of frozen sections and immunohistochemistry [2–4].

Major advantages of WSI are the possibility to analyse a slide from a remote access, share cases with experts and the inherent portability. In addition, WSIs enable visualisation of much more detail that the human eye is able to see by means of a conventional light microscope (CLM) [5]. WSI systems are more ergonomic, provide larger field of vision and easy navigation, allow a wider range of magnifications and make it possible to easily perform measurements and annotations [2]. These systems also provide high-quality digital images, which enable conservation of cases and may prevent loss of data. Cloud storages facilitate, eliminate storage problems, allow easy searching for case

✉ Alan Roger Santos-Silva
alan@unicamp.br

[1] Oral Diagnosis Department, Semiology and Oral Pathology Areas, Piracicaba Dental School, University of Campinas (UNICAMP), Av. Limeira, 901, Bairro Areião, Piracicaba, São Paulo 13414-903, Brazil

[2] Department of Oral Surgery and Pathology, School of Dentistry, Federal University of Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil

[3] Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, Sheffield, UK