The diagnostic accuracy of digital microscopy: a systematic review protocol

(Version 4)

Edward Goacher, Rebecca Randell and Darren Treanor

January 2015

Table of Contents

Background

Currently, histopathologists diagnose disease by analysing glass slides of tissue samples using a microscope. Advances in technology have made it possible to generate digital images of the same slides. Digital slides can then be viewed and assessed on a computer screen. Generated images are stored and shared virtually. Digital slides are currently used routinely in both education (both undergraduate and postgraduate) and research. Whilst digital slides are being used increasingly in some clinical settings, such as multi-disciplinary team (MDT) meetings and second opinions, they are not currently used in routine primary diagnosis.

There are multiple barriers to implementing the routine use of digital slides in a primary diagnostic setting. Such barriers include the high cost of implementation and a low acceptability amongst pathologists. However, the main barrier preventing their implementation is the lack of evidence validating its diagnostic accuracy in comparison to glass slides. It is necessary to determine whether digital slides can be considered to be an appropriate representation of the diagnostic information present on glass slides. This systematic review will review studies that compare the use of digital slides with the use of glass slides.

Aims

- Primary aim:
    - To assess the effect of digital microscopy on diagnostic accuracy.
- Secondary aims:
    - To examine the effect of digital slides on other elements of performance, such as speed of diagnosis and patient outcomes.
    - To identify which technologies (software and hardware) and study-level factors (e.g. case type, training provided) are associated with effective use of digital slides.

The objectives will be addressed using the following comparison:

Use of digital slides compared with use of glass slides.

To effectively evaluate the impact of current digital slide technologies, the review will prioritise those studies that compare the use of digital slides produced with whole slide imaging (WSI) with the use of glass slides. However, the number of such studies is likely to be low. Therefore the review will also include studies that compare the use of earlier technologies such as dynamic and static telepathology with the use of glass slides. This will allow a comparison of the diagnostic accuracy between digital slide technologies, as well as an examination into whether pathologist performance associated with the use of digital slides has changed as the technology has developed.

Study criteria

1.1 Types of studies

All studies comparing the diagnostic accuracy of digital slides and glass slides will be included and reviewed. Studies are likely to use the following designs:

- Crossover trials;
- Multiple reader multiple case studies; and
- Validation studies (where cases previously assessed using glass slides are reassessed.

1.2 Types of participants

Studies with the following participants will be included:

- Fully trained pathologists.

The type of participant from each study will be recorded. Any differences in the outcomes observed between different participants will also be recorded.

1.3 Types of intervention

Studies examining the following types of digital slide technologies will be considered:

- WSI

All studies comparing the use of digital slides vs. glass slides in the diagnostic process will be reviewed. The review will also consider any studies comparing the different types of digital slide technologies.

1.4 Types of performance and outcome measures

Our primary performance measure is diagnostic accuracy. An initial scan of the literature suggests that this is likely to be reported either as the mean number and standard deviation of correct diagnoses in the two conditions or as the rate of concordance of the diagnoses reached using the digital slides with the diagnoses reached using the glass slides.

In addition to diagnostic accuracy, we will abstract all reported performance and patient outcomes, categorising them into dichotomous outcomes and continuous outcomes, as well as an economic data that is reported.

Evaluating the effectiveness of digital slides is not a straightforward task. This is because digital slides are an example of a complex intervention, being made up of a number of components, including technological (the scanners and the hardware and software used for viewing the slides), clinical (the type of case/tissue) and organisational (the training provided, experience of using digital slides). Lack of improved performance in a particular outcome could be due to a variety of reasons, including access to software, inappropriate hardware, suboptimal use of the technology, or lack of time or support among colleagues. The systematic review will attempt to identify and reflect on such issues, particularly focusing on drawing out those types of cases where use of digital slides appears to be problematic.

Search strategy

1.5 Electronic databases to be searched

- Medline;
- Medline in progress;
- EMBASE; and

- Cochrane Central Register of controlled trials.

1.6 Alternate sources to be searched

- ClinicalTrials.gov;
- Reference searching of all papers and relevant reviews identified; and
- Contact Authors of relevant review papers and digital pathology experts regarding any further published or unpublished work.

Search strategies for electronic databases will be developed using selected MeSH terms ('telepathology') and free text terms generated from the PICO model:

Because the number of retrieved studies is expected to be small, no terms that attempt to limit the study design will be included, in order to keep the search as broad as possible. The search strategy will be translated into the other databases using the appropriate controlled vocabulary as applicable.

We will conduct a pilot search using the search strategy, iterating the process after identifying and incorporating additional keywords and text words used to describe and index the retrieved reports. Studies found only on the reference sections of the retrieved reports but missed by the search strategy will be

searched for and, if found in one of the electronic databases, its keywords will be added to the search strategy.

Citation tracking (both backwards and forwards) will be used to identify additional articles and alerts will be set up using the search criteria to inform us of new studies fitting criteria during the period of analysis.

1.7 Grey literature searching

To limit bias, a search of the existing grey literature will also be performed of the following sources:

- ProQuest Dissertations and Theses; and
- OVID HMIC (Health Care Management Information Consortium (DoH).

Review methods

1.8 Screening

Screening will be carried out using the algorithm that can be found in **Appendix A**. This algorithm is based on the criteria described above.

Two reviewers (Edward Goacher and Dr. Darren Treanor) will screen titles and abstracts for inclusion independently using the algorithm. Full text copies of all potentially eligible papers will be retrieved. The reproducibility of this process will be tested in the initial stages of the review and if reproducibility is shown to be poor, the algorithm will be refined to be more explicit. Where there is any uncertainty over whether or not a paper is eligible for inclusion, reviewers will discuss the papers and where possible resolve by consensus after referring to the protocol. When the reviewers are unable to decide between them, a third independent reviewer (Dr Rebecca Randell) will be consulted. Any disagreements and their resolution will be recorded.

All studies that initially appear to meet our inclusion criteria, but on closer inspection fail to, will be detailed in the table of excluded studies along with the reason for each exclusion. This table will be included in the report of the review as an appendix.

1.9 Data extraction

Two reviewers working independently will carry out data extraction on each paper. Data extraction will be carried out using the data extraction form that can be found in **Appendix B**. This data extraction form is based on the generic EPOC data collection template, modified to capture more detail in some areas. For all dichotomous outcomes, we will record the numbers in each of the two categories (event/no event) in each of the intervention groups. For continuous outcomes, we will record the mean values of the outcomes, the standard deviations of the outcomes, and the number of cases on whom the outcome was assessed in each of the two groups. We will write to authors to attempt to obtain important missing information.

The data extraction form will be converted into an electronic form, using Access. This electronic form will also be piloted. Once both reviewers are happy with the data extraction form, data extraction of all included studies will be carried out.

1.10 Quality assessment

The methodological quality of all included studies will be independently assessed by two reviewers using the QUADAS-2 tool.

As part of the data extraction form, the items in the QUADAS-2 tool have been assembled into a checklist, which can be used to systematically evaluate each study. Where necessary information is not available within the paper, the authors will be contacted for additional information.

The checklist will be piloted as part of the piloting of the data extraction sheet as a whole.

1.11 Data synthesis

For each study, we will report the main results in natural units in the results table. The studies will be categorised by system type (e.g. WSI, static telepathology, robotic microscopy). The following data will be included in the results table:

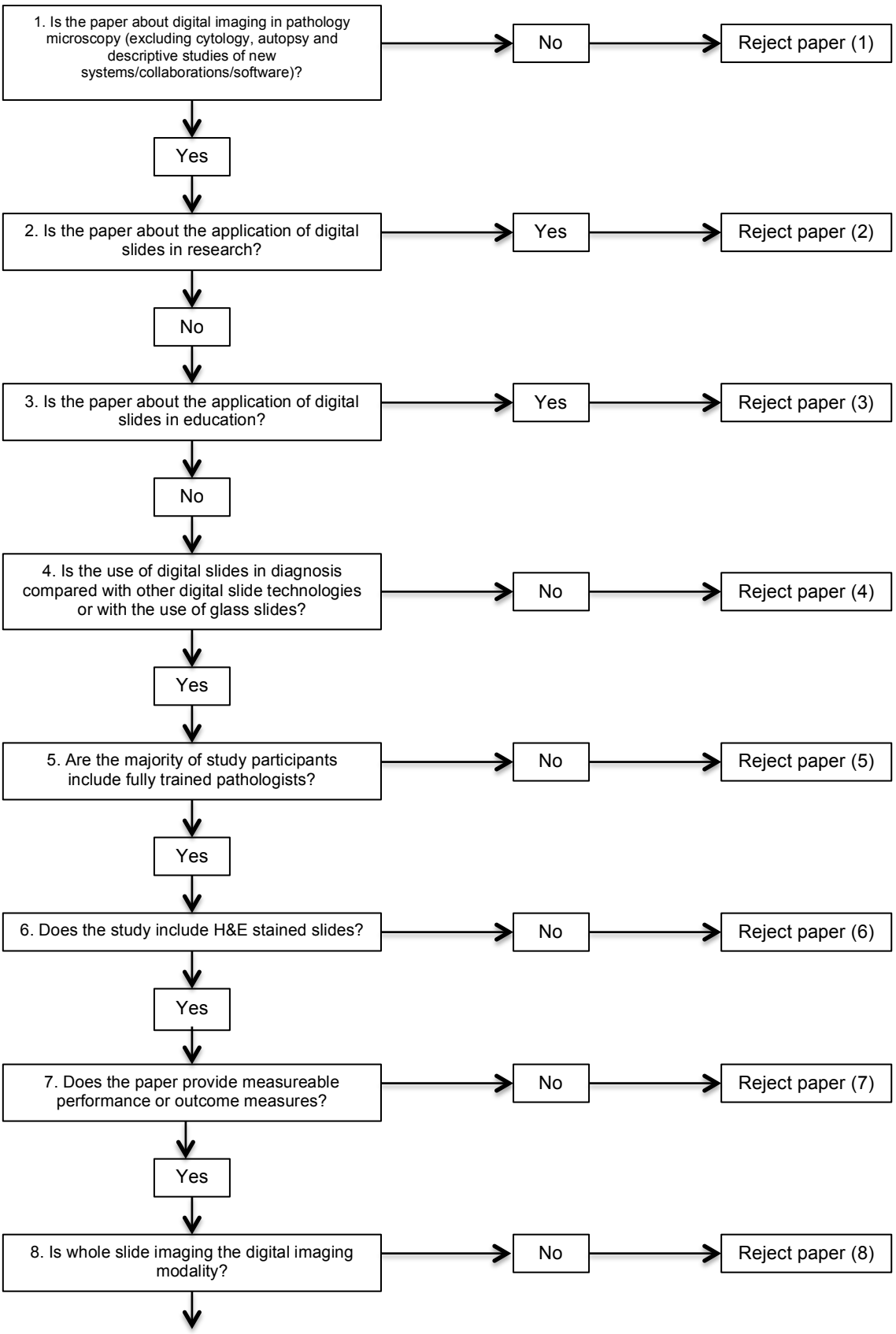
- System type
- Risk of bias
- Number of participants
- Number of cases
- Case type
- Diagnostic accuracy

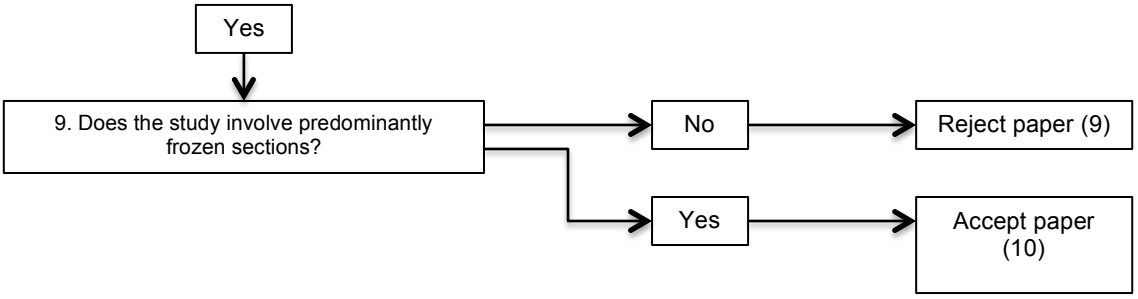
For meta-analysis to be justified the interventions should be the same and the outcome measures should be the same across studies. We will attempt to identify subgroups of studies with homogeneity in terms of study design, intervention type (e.g. WSI systems), and outcomes measured. Where such subgroups of studies are identified, we will carry out a random effects (DerSimonian and Laird) meta-analysis to calculate an average effect. Each study will be weighted according to its sample size and the resulting precision of the estimate of effect. For dichotomous data, meta-analysis will be performed using risk ratios and the result re-expressed as risk differences. For continuous data, standardised differences in means will be calculated.

Univariable and multivariable logistic regression models, adjusted for study risk of bias, will be used to investigate associations between outcomes of interest and study-specific covariates.

Data analysis will be performed using Microsoft Excel and Stata.

6. Appendix A


1. Is the paper about digital imaging in pathology microscopy (excluding cytology, autopsy and descriptive studies of new systems/collaborations/software)?
No
Reject paper (1)
Yes
2. Is the paper about the application of digital slides in research?
Yes
Reject paper (2)
No
3. Is the paper about the application of digital slides in education?
Yes
Reject paper (3)
No
4. Is the use of digital slides in diagnosis compared with other digital slide technologies or with the use of glass slides?
No
Reject paper (4)
Yes
5. Are the majority of study participants include fully trained pathologists?
No
Reject paper (5)
Yes
6. Does the study include H&E stained slides?
No
Reject paper (6)
Yes
7. Does the paper provide measureable performance or outcome measures?
No
Reject paper (7)
Yes
8. Is whole slide imaging the digital imaging modality?
No
Reject paper (8)



The number of studies rejected at stage 1 will be recorded, along with reasons why. Based on a previous literature review performed, it is likely that many of the studies rejected at stage 1 will be studies utilising digital slide technology for research purposes such as image analysis. The number of studies rejected at stage 2 will also be recorded, along with reasons why.

As discussed earlier, studies which initially appear to meet our inclusion criteria, but on closer inspection fail to, will be detailed in a table of excluded studies along with the reason for each exclusion.

7. Appendix B

Name of reviewer:

Date:

Study reference number:

1. Methods

1.1. Study design

1.2. Duration of study

1.3. First year pathologists were recruited to study

1.4. Last year pathologists were followed in study

1.5. Unit of allocation

1.6. Type of slide i.e. frozen section/paraffin section

1.7. Corresponding clinical details of slides provided

1.8. Washout period between modalities

1.9. Pathology specialty

1.10. Unit of analysis

1.11. Power calculation

1.12. Quality criteria (complete using QUADAS-2 guidance)

1.12.1. Patient selection

1.12.1.1. Risk of bias

1.12.1.1.1. Could the selection of patients have introduced bias?

1.12.1.1.2. Was a consecutive or random sample of patients involved?

1.12.1.1.3. Did the study avoid inappropriate exclusions?

1.12.2. Applicability

1.12.2.1. Are there concerns that the included patients and setting do not match the review question?

1.13. Index test

1.13.1. Risk of bias

1.13.1.1. Could the conduct or interpretation of the index test have introduced bias?

1.13.1.1.1. Were the index test results interpreted without the knowledge of the results of the reference standard?

1.13.1.1.2. Were the corresponding clinical details provided for each case?

1.13.1.1.3. Are participants trained in using the index test?

1.13.2. Applicability

1.13.2.1. Are there concerns that the index test, its conduct, or its interpretation differs from the review question?

1.14. Reference standard

1.14.1. Risk of bias

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

1.14.1.1.1. Is the reference standard likely to correctly classify the target condition?

1.14.1.1.2. Were the reference standard results interpreted without knowledge of the results of the index test?

1.14.1.1.3. Were the corresponding clinical details provided for each case?

1.14.2. Applicability

1.14.2.1.1. Are there concerns that the target condition as defined by the reference standard does not match the question?

1.15. Flow and timing

1.15.1. Risk of bias

1.15.1.1. Could the patient flow have introduced bias?

1.15.1.1.1. Was there an appropriate interval between the index test and the reference standard?

1.15.1.1.2. Did all patients receive the same reference standard?

1.15.1.1.3. Were all patients included in the analysis?

2. Participants

2.1. Number of histopathologists included in the study

2.2. Grade/qualifications of histopathologists

2.3. Specialty of histopathologists

2.4. Country

2.5. Environmental, social and cultural factors that may influence adherence

3. Interventions

3.1. Type of system:

- WSI

3.2. Scanner used

3.2.1. Make

3.2.2. Model

3.2.3. Mag

3.2.4. Compression type

3.2.5. Compression quality

3.3. Software used for viewing slides

3.4. Hardware used for viewing slides

3.5. Users were trained in use (Yes/No)

4. Outcomes

4.1. Outcome reported in what percentage of practitioners, units or clinics?

4.2. If follow-up was less than 100%, was there a description of withdrawals and dropouts?

4.3. Dichotomous outcome measures

- Intervention group: event
- Intervention group: no event
- Control group: event
- Control group: no event

4.4. Continuous outcome measures

- Intervention group: mean value of outcome
- Intervention group: standard deviation of outcome
- Intervention group: number of participants
- Control group: mean value of outcome
- Control group: standard deviation of outcome
- Control group: number of participan

Corresponding Author contact details

Mr. Edward W G Goacher

88 Meaford Road,
Barlaston,
Staffordshire,
England,
ST12 9EB.

Phone: 07854438101
Email: um11eg@leeds.ac.uk