BMJ Open

# The CONFIDENT-trial protocol: a pragmatic template for clinical implementation of artificial intelligence assistance in pathology



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2022-067437 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 17-Aug-2022 |
| Complete List of Authors: | Flach, Rachel; University Medical Centre Utrecht, Department of Pathology<br>Stathonikos, Nikolas; University Medical Centre Utrecht, Department of Pathology<br>Nguyen, Tri; University Medical Centre Utrecht, Department of Pathology<br>Ter Hoeve, Natalie; University Medical Centre Utrecht, Department of Pathology<br>van Diest, P.J.; University Medical Centre Utrecht, Department of Pathology<br>van Dooijeweert, Carmen; University Medical Centre Utrecht, Department of Pathology |
| Keywords: | Prostate disease < UROLOGY, Breast tumours < ONCOLOGY, PATHOLOGY |
| | |

# The CONFIDENT-trial protocol: a pragmatic template for clinical implementation of artificial intelligence assistance in pathology


## Authors

R.N. Flach (r.n.flach-2@umcutrecht.nl), N. Stathonikos (nstatho2@umcutrecht.nl), T.Q. Nguyen (t.q.nguyen@umcutrecht.nl), N.D. ter Hoeve (N.terhoeve@umcutrecht.nl), P.J. van Diest (P.j.vandiest@umcutrecht.nl), C. van Dooijeweert (c.vandooijeweert3@umcutrecht.nl)

## Affiliation

Department of Pathology, University Medical Centre Utrecht, Utrecht, The Netherlands


## Trial registration

Applicable nor suitable.

## Protocol version

Version 1.


## Funding

No funding was received for this research.


## Roles and responsibilities

As no funding was received, and as this is a single centre clinical trial, full authority over all study activities (design; collection, management, analysis and interpretation of data; writing of the report; and the decision to submit the report for publication) lies with the researchers in the UMC Utrecht.


## Corresponding author

Carmen van Dooijeweert

Internal mail no H04.312

T: +31 88 75 565 65

c.vandooijeweert3@umcutrecht.nl

# Abstract

## Introduction

Artificial Intelligence (AI) has been on the rise in the field of pathology. Despite promising results in retrospective studies, and several CE-IVD certified algorithms on the market, prospective clinical implementation studies of AI have yet to be performed. In this trial, we will explore the benefits of an AI-assisted pathology workflow, while maintaining diagnostic safety standards.

## Methods and analysis

This is a SPIRIT-AI compliant single-centre, controlled clinical trial, in a fully digital academic pathology laboratory. We will prospectively include prostate cancer patients who undergo prostate needle biopsies (CONFIDENT-P) and breast cancer patients who undergo a sentinel lymph node (SN) procedure (CONFIDENT-B) in the UMC Utrecht. For both the CONFIDENT-B and CONFIDENT-P trials, the specific pathology specimens will be pseudo-randomized to be assessed by a pathologist with- or without AI-assistance in a pragmatic (bi-)weekly sequential design. In the intervention group, pathologists will assess whole slide images (WSI) of the standard haematoxylin-eosin (HE)-stained sections assisted by the output of the algorithm. In the control group, pathologists will assess HE WSI according to the current clinical workflow. If no tumour cells are identified or when the pathologist is in doubt, staining by immunohistochemistry (IHC) will respectively may be performed. At least 80 patients in the CONFIDENT-P and 180 patients in the CONFIDENT-B trial will need to be enrolled to detect superiority, allocated as 1:1. Primary endpoint for both trials is the number of saved resources on IHC for detecting tumour cells, since this will clarify tangible cost savings that will build the business case for AI.

## Ethics and dissemination

The ethics committee (MREC NedMec) waived the need of official ethical approval, as participants are not subjected to procedures and as they are not required to follow rules. Results of both trials (CONFIDENT-B and CONFIDENT-P) will be published in scientific peer-reviewed journals.


# Strengths and limitations

- This is the first clinical trial to examine the added value of artificial intelligence in the daily pathology workflow.
- By maintaining the current diagnostic safety standards patients are not at risk of an inferior diagnosis during the trial.
- This is a pragmatic template for prospective AI-trials for object-identifying algorithms in pathology.
- A limitation is that this is a single-centre study, which may hamper generalizability.
- Due to the existing clinical workflow, randomization of patients and (double-)blinding of the participating pathologists and researchers is not possible.

**Keywords:**
Artificial intelligence, prospective trial, prostate cancer, breast cancer, digital pathology

Word count: 2864 words

# Introduction

## Background and rationale

Since the introduction of digital pathology, the number of studies on artificial intelligence (AI) within the field of pathology has increased exponentially.[1,2] Algorithms have been created for tumour detection, tumour grading, recognizing tumour subtypes, evaluating biomarkers and more.[1,3] Worldwide, a shortage of pathologists exists, while their workload is only increasing.[2,4] Therefore, AI has great potential to alleviate pathologists' workload.[2] At the same time, AI has great potential to improve diagnostics by improving accuracy, reproducibility and speed.[2] In fact, several algorithms have shown to be comparable, or even superior to pathologists (under time-constraint).[2,5–10] However, artificial and human intelligence are not mutually exclusive, but they complement each other, a concept which is known as "augmented intelligence", where AI enhances, rather than replaces human intelligence.[11] In the (very) early AI-adoption phase, and presumably also in later adoption phases, pathologist-supervision remains of key-importance.

This is particularly relevant as, despite the promising results of retrospective studies and the availability of CE-IVD approved algorithms, prospective validation and clinical implementation of AI are currently lacking. For example, six years after the successful CAMELYON-16 Grand Challenge[6], the top algorithms have yet to be implemented in daily clinical practice, showing that the time between development of an AI model and clinical implementation is considerable. Likewise, numerous promising prostate cancer grading algorithms have been developed, yet implementation studies have yet to be performed.[12–14] In addition, nine AI pathology devices received CE-IVD approval in 2021.[15]

## Trial rationale

As a pathology laboratory with a fully digital workflow for over seven years, we are eager to explore the full potential of working digitally by adding the benefit of AI in daily pathology practice. We decided to start with the object localisation task of tumour-detection, where an objective reference standard is in place in the routine clinical workflow (i.e. pathologist-supervision and/or immunohistochemistry (IHC) staining).

We developed the CONFIDENT-trial template, in which AI tumour detection algorithms can be safely implemented in prospective clinical trials, while ensuring that patients are not at risk of receiving an inferior diagnosis, since IHC is always performed when no tumour cells are visible, but also when pathologists need more confirmation about the diagnosis.

## CONFIDENT-B and CONFIDENT-P

Our trials aim to prospectively investigate the added value of an AI-assisted pathology workflow in the identification of prostate cancer (PCa) in prostate needle biopsies (CONFIDENT-P) and the identification of sentinel lymph node (SN) metastases in breast cancer (BCa) patients (CONFIDENT-B). As both prostate cancer and breast cancer are the most common (non-skin) malignancies in men and women, respectively, implementation of AI-assistance may have a great impact on diagnostic processes.[16] However, it is important to emphasize that this trial serves as a template for other pragmatic AI-intervention trials for object-localisation tasks as well.

We obtained CE-IVD-approved algorithms for detection and grading of prostate cancer in prostate needle biopsies and an algorithm for detecting lymph node metastases in BCa patients. In

both cases the task of the pathologist is both labour-intensive and expensive, due to the performed IHC stains in case no tumour cells are morphologically observed. However, IHC is expensive and these costs sometimes even exceed reimbursement for the entire specimen (e.g. in case of multiple blocks of multiple SNs). This raises the question whether AI may be of added value to morphologically detect cancer cells without the need for IHC-use. Thereby the number of performed IHC stains may be reduced, which may lead to tangible costs savings that will help to build the business case for AI, while potentially decreasing the workload of pathologists as well.[2]

### Study objective

The primary objective is to explore whether an AI-assisted workflow reduces the number of spent resources on IHC, while maintaining diagnostic safety standards in both PCa patients who underwent prostate needle biopsies (CONFIDENT-P), and BCa patients who underwent an SN procedure.

Secondary objectives are to investigate whether time management improves in an AI-assisted workflow and to analyse how many IHC staining may have been safely omitted after AI-implementation.

## Methods and analysis

### Trial design

The study protocol is structured following the SPIRIT-AI (Standard Protocol Items: recommendations for Interventional Trials - Artificial Intelligence) statement 2020[17]. This study is a single-centre, parallel-group controlled trial, assessing superiority. The allocation ratio is 1:1. Eligible patients will be assigned to arm 1 (control group) or arm 2 (AI-assisted workflow), based on a bi-weekly time-schedule. Eligibility criteria are summarised in Figure 1. The CONFIDENT trials will be carried out in 2022-2023.

### Study setting

The trial will take place in the daily practice of a single academic hospital (University Medical Centre Utrecht, the Netherlands), with a fully digital pathology enabled clinical set-up, where all slides are digitised using ultrafast whole slide image (WSI) using Hamamatsu S360 scanners and reviewed using the Sectra pathology Picture Archiving and Communication System (PACS). Although the UMC Utrecht is an academic hospital, primary routine pathology diagnostics is performed for non-academic hospitals as well (i.e., the Alexander Monro Breast Cancer Hospital, Bilthoven, the Netherlands).

### Study population

For PCa, WSI of all males who undergo a prostate needle biopsy in the UMC Utrecht will be included. For BCa, WSI of all females or males with BCa as primary malignancy (i.e. invasive breast cancer) who undergo an SN procedure in the Alexander Monro Breast Cancer Hospital or the UMC Utrecht will be included. Patients will be excluded, if they were redirected to the UMC Utrecht for a second opinion.

### Assessment of specimen

During the study period, all WSI will be assessed by the same group of pathologists; i.e. two expert urological pathologists for the PCa biopsies, and three expert breast pathologists for the lymph node assessment from BCa patients.

For both the CONFIDENT-B and CONFIDENT-P trial, the specific pathology specimens will be assigned to be assessed by a pathologist with or without AI-assistance in a pragmatic (bi-)weekly sequential design. This is deemed feasible as case-mix variation and time trends are highly unlikely to occur within the envisioned inclusion-period of approximately six to nine months. Furthermore, both

specialized breast- and urological pathologists within the UMC Utrecht work according to weekly schedules. Therefore, using AI every other week or every other two weeks, as opposed to switching by day, ensures that all pathologists are equally distributed between groups. Lastly, it would be impractical to switch from AI-assistance to no AI-assistance on a case-to-case basis. For obvious reasons, allocation concealment and blinding of pathologists is not applicable. In addition, as there is no room for interpretation, researchers who perform the data-analysis will also not be blinded.

## Control and intervention

All eligible specimens will be assigned to either the control group or the intervention group. In the control group, pathologists will assess H&E stained WSI of patients digitally, according to the current clinical workflow. For PCa biopsies, IHC is routinely performed on all cases. For BCa lymph nodes, if no metastases or tumour are present, IHC staining will be performed. Additional IHC staining will also be performed by additional request of the pathologist in case of doubt.

In the intervention group, pathologists will assess the H&E-specimens digitally with the outcome of the algorithm provided in their first assessment of the specimen. For PCa, they will use the CE-IVD certified Paige Prostate Suite algorithm, and for BCa, pathologists will use the CE-IVD certified Metastasis Detection App by Visiopharm. These algorithms will be integrated within the Sectra PACS where the output of the algorithms will be graphically displayed. AI analysis of the WSI will be performed right after scanning to avoid delays in the clinical workflow. If the AI-assisted pathologist does not detect metastases or tumours on the H&E slide, routine additional IHC staining will be performed by P503S/p63/CK HMW for PCa and CAM5.2 for BCa, to ensure no metastases or tumours are missed. Pathologists can also request an additional IHC if they feel they need this to make an adequate diagnosis (Figure 2).

## Outcome measures

Primary outcome for the CONFIDENT-P trial is the added value of AI-assistance in the detection of PCa and in the detection of tumour volume in prostate needle biopsies in daily pathology practice. The primary outcome for the CONFIDENT-B trial is the added value of AI-assistance in the detection of BCa SN-metastases. The outcome measures for both trials will be the number of spent resources, i.e. the number of IHC-stains performed in both groups.

Secondary outcome measures will be sensitivity and specificity of the AI-assisted pathologist, time spent on WSI analysis, the number of IHC stains that may have been omitted after AI-implementation and a pathologists' evaluation by a questionnaire om the AI-assisted work process.

## Input data

Input data for the algorithm will be WSI of haematoxylin and eosin (H&E) stained slides scanned at 40x of either prostate needle biopsies, and WSI of H&E stained slides of BCa lymph nodes. As per routine in our daily clinical practice, WSI will be quality controlled after scanning for colour, focus quality and completeness of the scan. When necessary, the specimens will be rescanned.

## Sample size

**CONFIDENT-P**

We performed power calculations using a two-sample proportion superiority test, using expected percentages of IHC staining in both study arms. We assume that the pathologists in the control arm can detect 50% of the tumours without using IHC. We expect AI-assisted pathologists to detect 80% of the tumours, without using IHC. These percentages were conservatively derived from the validity

study by Raciti (74% for pathologists without AI and 90% for pathologists with AI respectively)[18], by expert pathologist opinion, and taking into account that pathologists under time constraint of daily practice do not detect tumours as well as pathologists without time constraint during retrospective studies[19]. We assume that this effect will be larger for the biopsies assessed without AI than with AI, as AI is assumed to make tumour detection easier.

A sample size of 60 (30 per arm) would give a power of approximately 80%, using a one-sided 5% significance level. However, uncertainties remain regarding the sample size parameters. We therefore inflated our sample size to 80 (40 per arm), in order to ensure study power and allowing us to detect smaller effect sizes.

For detection of tumour volume percentage, we performed a power calculation based on the assumption that AI should be able to replace at least 20% of the IHC stains, in order to be cost-effective. IHC is currently used in 100% of all prostate needle biopsies. Using a power of 80% and a one-sided significance level of 5%, this leads to 27 patients per arm.

**CONFIDENT-B**

Sample size calculations for the CONFIDENT-B trial are based on the assumption that the AI-algorithm can detect all metastases for which currently IHC is used, which are mainly micro-metastases and isolated tumour cells (ITC). Approximately 15% of the SN-specimens in the UMC Utrecht contain a micrometastasis or ITC.

A sample size of 166 patients (83 per arm) with a one-sided 5% significance level therefore results in a power of 80%. Again, as there are uncertainties on the assumptions on what amount of the metastases will be detect by AI, we decided to be conservative and include 180 patients (90 per arm).

Overall, we are only interested in one-sided outcomes, as it is not possible that more IHC will be performed in the AI-assisted arm. IHC is performed to detect metastases, when they are macroscopically undetectable, rather than to confirm them when they are macroscopically visible. As AI would show only more metastases than the pathologist could macroscopically detect, this means that only a reduction of IHC is possible.

## Statistical methods

For baseline comparisons between both arms, the appropriate measures (parametric or non-parametric) for categorical (Chi-squared test/Fisher's exact) and continuous variables (T-test, Mann-Whitney U test) will be used. For the analysis of the primary outcome measure, we will compare the proportion of IHC-use in both arms, and calculate adjusted relative risks, using a log-binomial model.[20–22]

Missing data for baseline characteristics and for the primary outcome are not to be expected, as they are obligatory items in the structured pathology reports.

We will determine sensitivity and specificity of the conclusion of the AI-assisted pathologists without the use of IHC. Subsequently, we will focus on the cases with metastases that the AI-assisted pathologist misses, categorize them (i.e. macro-metastases, micro-metastases, ITC) and determine their clinical relevance (i.e. clinical consequences if these metastases are being missed). Data-analysis will be performed in R Statistical Software[23], with a significance level set at $p<0.05$.

## Data collection and management

All data (baseline and primary outcome measurements) will be retrieved from the structured pathology reports and will be managed and stored in Castor EDC[24]. For the secondary outcome measure of time spent by the pathologist on a slide, data will be collected on an interval basis for

practical reasons, as timing every assessment for months (by stopwatch) was not deemed feasible. For the secondary outcome measure of AI-assisted work process for pathologists, a questionnaire will be distributed to the participating pathologists. The final secondary assessment measurement, the number of IHC stains that may have been omitted after AI-implementation, will be determined by the researchers based on the data from the structured pathology reports (combination of IHC and AI-assisted diagnoses of the pathologist).

### Ethical approval

Research within these trials is not subject to the (Dutch) Medical Research Involving Human Subjects Act (WMO), as participants are not subjected to procedures and as they are not required to follow rules. Therefore, the ethics committee (MREC NedMec) waived the need of ethical approval and informed consent.

### Risk of harm

Patients are not at risk of any harm for an inferior diagnosis (i.e. missed tumour cells), as in both arms, IHC-staining will be performed when no tumour cells are visible, according to current clinical standards. As a rule in augmented intelligence, all cases will be evaluated by a pathologist, which further minimized the risk of a false diagnosis based on the AI algorithm. Taking all of the above into account, a data monitoring committee (DMC) is not required and adverse events are not to be expected. In theory, the algorithm could be more of a disturbance than a help to pathologists (for example, when it frequently reports false positive or negative results, which have to be corrected by the pathologist). However, the algorithms used are IVDR-approved, and thus have undergone extensive review for their intended purpose. Nonetheless, the experience and ease of use of pathologists working with the algorithm will be one of the secondary outcome measures.

### Informed consent and data access

Informed consent was waived by the local quality coordinator (QC) and data protection officer (DPO) for the following reasons. First, in both arms patients receive standard care, while maintaining diagnostic safety standards (pathologists' supervision, IHC in all negative cases). Second, patients are not subjected to any procedures. Third, all data will be anonymized to the researchers by the pathologist who assessed the slide.

The collected (anonymous) research data will be stored in Castor EDC to ensure data security. Data will be kept for a period of 15 years. Data access in Castor will be restricted to two researchers (RF, CvD). Pathologists have access to the electronic patient files for the purpose of patient care. The researchers are not permitted access to these files. At no point will the data (both in Castor EDC and patient files) be accessed by the companies providing the algorithms (i.e. Visiopharm and Paige).

### Patient and Public Involvement

None

## Discussion

The promising retrospective results of AI-assisted pathology have not yet resulted in prospective clinical implementation studies. This may be due to a lack of digital transition in the majority of pathology laboratories, but it may also be partly due to the lack of a good implementation model. Fortunately, however, new guidelines for AI-trials have recently been proposed by the SPIRIT-AI and CONSORT-AI-steering groups, as well as roadmaps to routine use of AI in clinical practice.[17,25,26] Yet, to date, no pathology AI-trials have been published in PubMed or Web of Science or, to the best of

our knowledge, otherwise made public.

As a pathology laboratory with a fully digital workflow, we developed a clinical trial template for tumour detection models, as a first step to implement AI in daily pathology practice. We will start with an object localisation task (i.e. tumour cells) as a reference standard is in place in the routine clinical workflow. For classification tasks like tumour grading, a clinical trial design is more challenging, as no reference is in place in daily pathology practice and inter-laboratory and inter-pathologist variation is notorious.[27–31] Nevertheless, in future trials, implementing AI-assistance in the grading process might also reduce this variation. For now, results of the CONFIDENT-trials will provide the first assessment of the potential added value of AI in daily pathology practice. This evaluation will substantially contribute to a potential paradigm shift in tumour detection in pathology. The pragmatic template of the CONFIDENT trials may serve as example for other prospective AI implementation trials in diagnostic pathology.

## Declaration of interests

PJvD is a member of the Scientific Advisory Board of Paige and Sectra.
All other authors do not report conflict of interest.

## Funding

No funding was obtained at the moment of writing this paper.

## Dissemination policy

Results of both trials (CONFIDENT-B and CONFIDENT-P) will be published in scientific peer-reviewed journals. Authorship will be acknowledged to all those that substantially contributed in the CONFIDENT-trials. Data will be available upon reasonable request.

## Author statement

PJvD conceived of the study. RNF and CvD initiated the study design and NS, TQN and NDtH helped with implementation. RNF and CvD provided statistical expertise in clinical trial design and RNF and CvD are conducting the primary statistical analysis. All authors contributed to refinement of the study protocol and approved the final manuscript.

## Figure legends

Figure 1. Flowchart with patient selection; SN = sentinel node; AI = artificial intelligence.

Figure 2. Study flow chart. H&E = hematoxylin & eosin; IHC = immunohistochemistry

## References


1. Jiang, Y., Yang, M., Wang, S., Li, X. & Sun, Y. Emerging role of deep learning-based artificial intelligence in tumour pathology. *Cancer Commun.* **40**, 154–166 (2020).
2. van der Laak, J., Litjens, G. & Ciompi, F. Deep learning in histopathology: the path to the clinic. *Nat. Med.* **27**, 775–784 (2021).
3. Steiner, D. F., Chen, P. H. C. & Mermel, C. H. Closing the translation gap: AI applications in digital pathology. *Biochim. Biophys. Acta - Rev. Cancer* **1875**, 188452 (2021).
4. Robboy, S. J. *et al.* Pathologist workforce in the United States: I. Development of a predictive model to examine factors influencing supply. *Arch. Pathol. Lab. Med.* **137**, 1723–1732 (2013).
5. Bulten, W. *et al.* Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet. Oncol.* **21**, 233–241 (2020).
6. Ehteshami Bejnordi, B. *et al.* Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* **318**, 2199–2210 (2017).
7. Steiner, D. F. *et al.* Evaluation of the Use of Combined Artificial Intelligence and Pathologist Assessment to Review and Grade Prostate Biopsies. *JAMA Netw. open* **3**, e2023267 (2020).
8. Ström, P. *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet. Oncol.* **21**, 222–232 (2020).
9. Raciti, P. *et al.* Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Mod. Pathol.* **33**, 2058–2066 (2020).
10. Hekler, A. *et al.* Superior skin cancer classification by the combination of human and artificial intelligence. *Eur. J. Cancer* **120**, 114–121 (2019).
11. Harrison, J. H. *et al.* Introduction to Artificial Intelligence and Machine Learning for Pathology. *Arch. Pathol. Lab. Med.* **145**, 1228–1254 (2021).
12. Ström, P. *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol.* **21**, 222–232 (2020).
13. Steiner, D. F. *et al.* Evaluation of the Use of Combined Artificial Intelligence and Pathologist Assessment to Review and Grade Prostate Biopsies. *JAMA Netw. Open* **3**, 1–14 (2020).
14. Bulten, W. *et al.* Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. *Mod. Pathol.* **34**, 660–671 (2021).
15. Muehlematter, U. J., Daniore, P. & Vokinger, K. N. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit. Heal.* **3**, e195–e203 (2021).
16. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
17. Cruz Rivera, S. *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *The Lancet Digital Health* vol. 2 e549–e560 at https://doi.org/10.1016/S2589-7500(20)30219-3 (2020).
18. Raciti, P. *et al.* Novel artificial intelligence system increases the detection of prostate cancer in

whole slide images of core needle biopsies. *Mod. Pathol.* **33**, 2058–2066 (2020).

19. Bejnordi, B. E. *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA - J. Am. Med. Assoc.* **318**, 2199–2210 (2017).

20. McNutt, L. A., Wu, C., Xue, X. & Hafner, J. P. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am. J. Epidemiol.* **157**, 940–943 (2003).

21. Lumley, T., Kronmal, R. & Ma, S. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. . *UW Biostat. Work. Pap. Ser. 293 [http//www. bepress. com/uwbiostat/paper293].* (2006).

22. Zou, G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am. J. Epidemiol.* **159**, 702–706 (2004).

23. R Core Team. R: A Language and Environment for Statistical Computing. at https://www.r-project.org/ (2020).

24. Castor Electronic Data Capture. at https://castoredc.com (2019).

25. Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J. & Denniston, A. K. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet. Digit. Heal.* **2**, e537–e548 (2020).

26. Colling, R. *et al.* Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J. Pathol. J Pathol* **249**, 143–150 (2019).

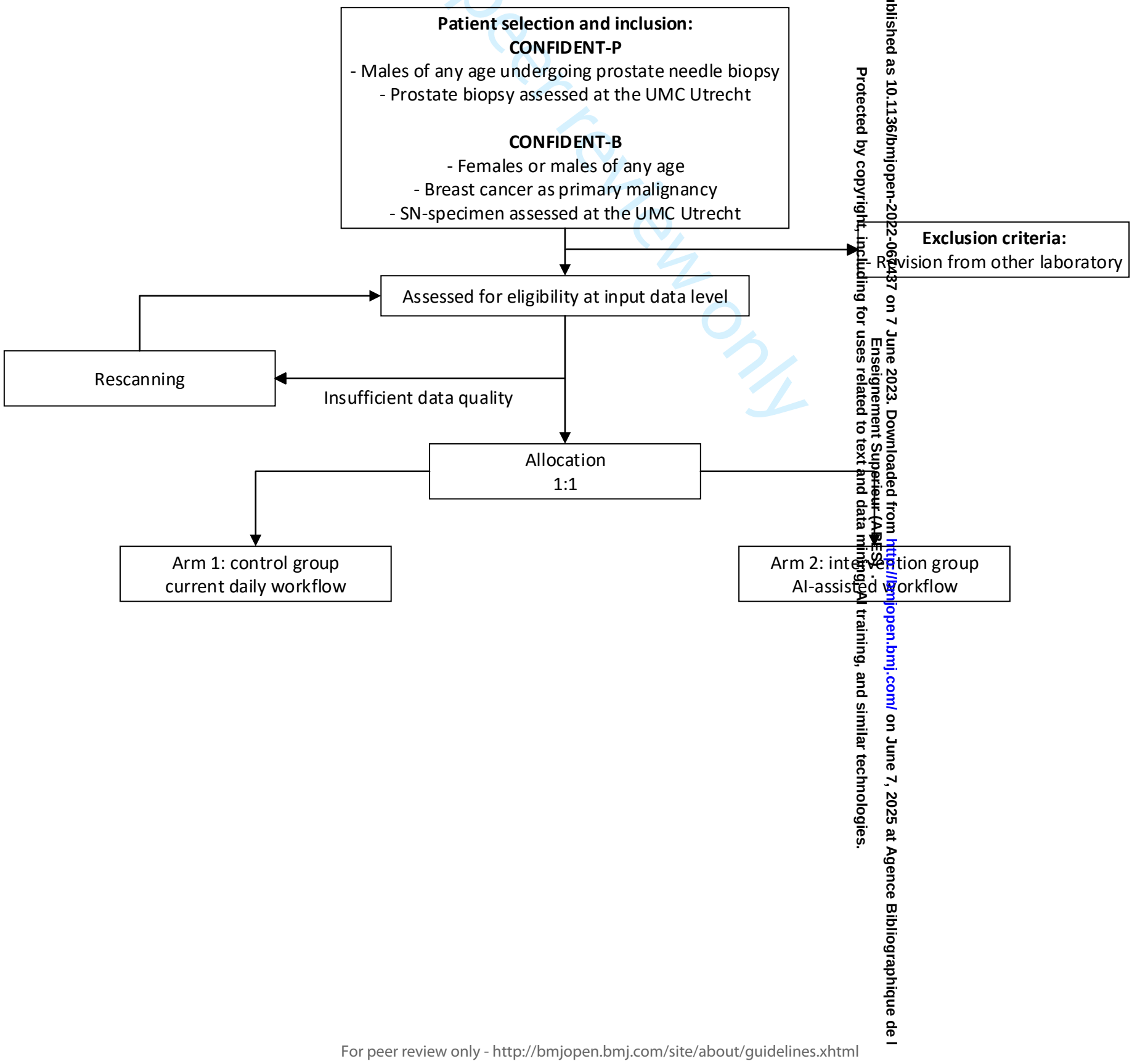
27. Kuijpers, C. C. H. J. *et al.* Interlaboratory variability in the histologic grading of colorectal adenocarcinomas in a nationwide cohort. *Am. J. Surg. Pathol.* **40**, 1100–1108 (2016).

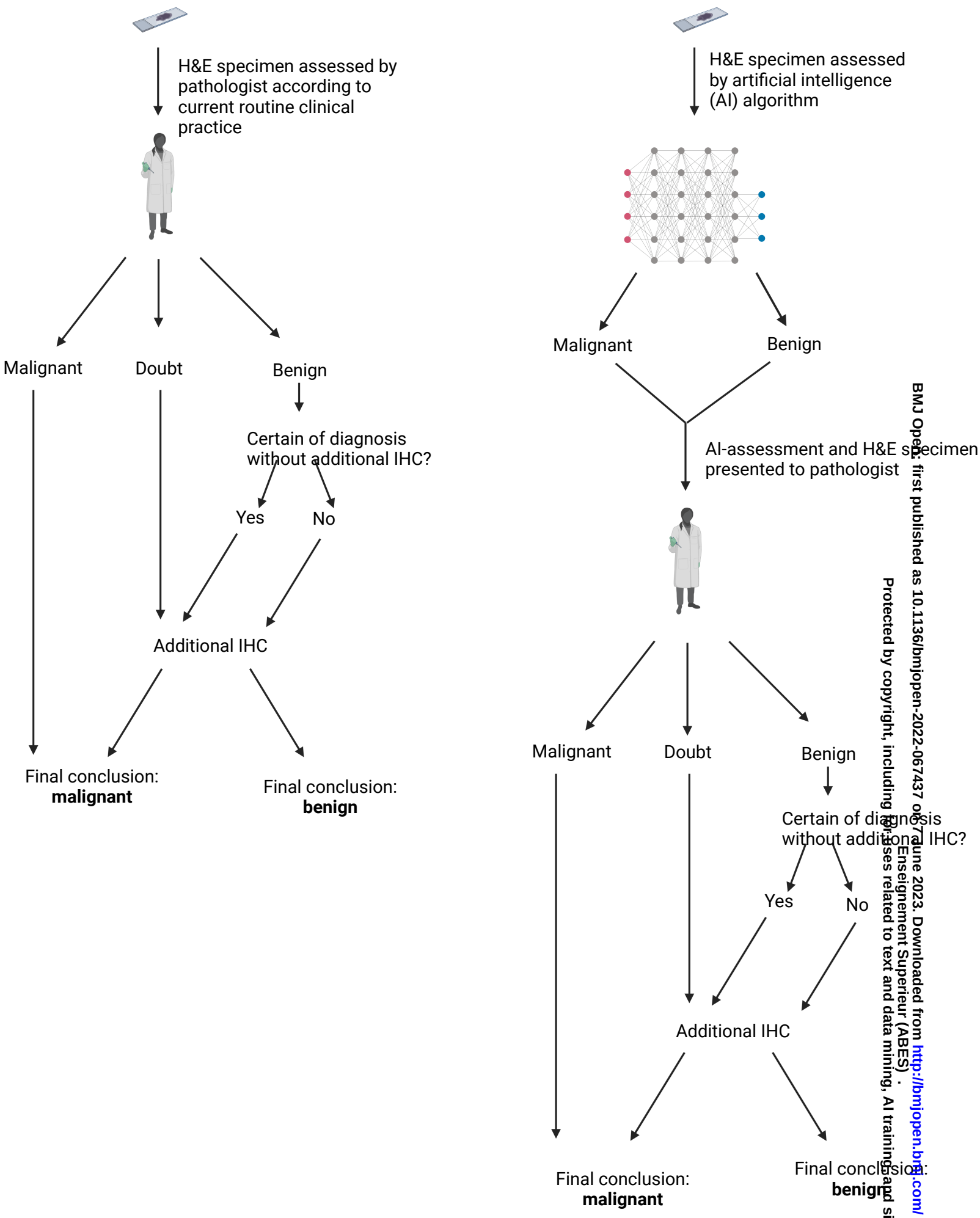
28. Kuijpers, C. C. H. J. *et al.* Interlaboratory variability in the grading of dysplasia in a nationwide cohort of colorectal adenomas. *Histopathology* **69**, 187–197 (2016).

29. Flach, R. N. *et al.* Significant Inter- and Intralaboratory Variation in Gleason Grading of Prostate Cancer: A Nationwide Study of 35,258 Patients in The Netherlands. *Cancers (Basel).* **13**, (2021).

30. van Dooijeweert, C. *et al.* Significant inter- and intra-laboratory variation in grading of ductal carcinoma in situ of the breast: a nationwide study of 4901 patients in the Netherlands. *Breast Cancer Res. Treat.* **174**, 479–488 (2019).

31. van Dooijeweert, C. *et al.* Significant inter- and intra-laboratory variation in grading of invasive breast cancer: A nationwide study of 33,043 patients in the Netherlands. *Int. J. Cancer* **146**, 769–780 (2020).





# Reporting checklist for protocol of a clinical trial.

Based on the SPIRIT guidelines.

## Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the SPIRITreporting guidelines, and cite them as:

Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin J, Dickersin K, Hróbjartsson A, Schulz KF, Parulekar WR, Krleža-Jerić K, Laupacis A, Moher D. SPIRIT 2013 Explanation and Elaboration: Guidance for protocols of clinical trials. BMJ. 2013;346:e7586

| | | Reporting Item | Page Number |
|---|---|---|---|
| **Administrative information** | | | |
| Title | #1 | Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym | 1 |
| Trial registration | #2a | Trial identifier and registry name. If not yet registered, name of intended registry | na |
| Trial registration: data set | #2b | All items from the World Health Organization Trial Registration Data Set | na |
| Protocol version | #3 | Date and version identifier | 1 |
| Funding | #4 | Sources and types of financial, material, and other support | 1 |
| Roles and responsibilities: contributorship | #5a | Names, affiliations, and roles of protocol contributors | 1 |

| | | | |
|---|---|---|---|
| Roles and responsibilities: sponsor contact information | #5b | Name and contact information for the trial sponsor | na |
| Roles and responsibilities: sponsor and funder | #5c | Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities | na |
| Roles and responsibilities: committees | #5d | Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) | na |
| **Introduction** | | | |
| Background and rationale | #6a | Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention | 3 |
| Background and rationale: choice of comparators | #6b | Explanation for choice of comparators | 3 |
| Objectives | #7 | Specific objectives or hypotheses | 4 |
| Trial design | #8 | Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, non-inferiority, exploratory) | 4 |
| **Methods: Participants, interventions, and outcomes** | | | |
| Study setting | #9 | Description of study settings (eg, community clinic, academic hospital) and list of countries where data will | 4-5 |

| | | | |
|---|---|---|---|
| | | be collected. Reference to where list of study sites can be obtained | |
| Eligibility criteria | #10 | Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) | 5 |
| Interventions: description | #11a | Interventions for each group with sufficient detail to allow replication, including how and when they will be administered | 6 |
| Interventions: modifications | #11b | Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving / worsening disease) | na |
| Interventions: adherance | #11c | Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return; laboratory tests) | na |
| Interventions: concomitant care | #11d | Relevant concomitant care and interventions that are permitted or prohibited during the trial | na |
| Outcomes | #12 | Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended | 6, 8 |
| Participant timeline | #13 | Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure) | 6 |
| Sample size | #14 | Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations | 7 |
| Recruitment | #15 | Strategies for achieving adequate participant enrolment to reach target sample size | 5 |

| | | | |
|---|---|---|---|
| **Methods: Assignment of interventions (for controlled trials)** | | | |
| Allocation: sequence generation | #16a | Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions | 5 |
| Allocation concealment mechanism | #16b | Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned | na |
| Allocation: implementation | #16c | Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions | na |
| Blinding (masking) | #17a | Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how | na |
| Blinding (masking): emergency unblinding | #17b | If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial | na |
| **Methods: Data collection, management, and analysis** | | | |
| Data collection plan | #18a | Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. | 8 |

| | | | |
|---|---|---|---|
| | | Reference to where data collection forms can be found, if not in the protocol | |
| Data collection plan: retention | #18b | Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols | na |
| Data management | #19 | Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol | 8-9 |
| Statistics: outcomes | #20a | Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol | 7-8 |
| Statistics: additional analyses | #20b | Methods for any additional analyses (eg, subgroup and adjusted analyses) | 7-8 |
| Statistics: analysis population and missing data | #20c | Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) | 7-8 |
| **Methods: Monitoring** | | | |
| Data monitoring: formal committee | #21a | Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed | 8 |
| Data monitoring: interim analysis | #21b | Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial | na |
| Harms | #22 | Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events | 8 |

| | | | |
|---|---|---|---|
| | | and other unintended effects of trial interventions or trial conduct | |
| Auditing | #23 | Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor | na |
| **Ethics and dissemination** | | | |
| Research ethics approval | #24 | Plans for seeking research ethics committee / institutional review board (REC / IRB) approval | 9 |
| Protocol amendments | #25 | Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC / IRBs, trial participants, trial registries, journals, regulators) | na |
| Consent or assent | #26a | Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) | na |
| Consent or assent: ancillary studies | #26b | Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable | 8 |
| Confidentiality | #27 | How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial | 8 |
| Declaration of interests | #28 | Financial and other competing interests for principal investigators for the overall trial and each study site | 9 |
| Data access | #29 | Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators | 8-9 |
| Ancillary and post trial care | #30 | Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation | na |
| Dissemination policy: trial results | #31a | Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the | 9 |

| | | | |
|---|---|---|---|
| | | public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions | |
| Dissemination policy: authorship | #31b | Authorship eligibility guidelines and any intended use of professional writers | 9 |
| Dissemination policy: reproducible research | #31c | Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code | 9 |
| **Appendices** | | | |
| Informed consent materials | #32 | Model consent form and other related documentation given to participants and authorised surrogates | 8 |
| Biological specimens | #33 | Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable | na |

BMJ Open

# The CONFIDENT-trial protocol: a pragmatic template for clinical implementation of artificial intelligence assistance in pathology



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2022-067437.R1 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 22-Mar-2023 |
| Complete List of Authors: | Flach, Rachel; University Medical Centre Utrecht, Department of Pathology<br>Stathonikos, Nikolas; University Medical Centre Utrecht, Department of Pathology<br>Nguyen, Tri; University Medical Centre Utrecht, Department of Pathology<br>Ter Hoeve, Natalie; University Medical Centre Utrecht, Department of Pathology<br>van Diest, P.J.; University Medical Centre Utrecht, Department of Pathology<br>van Dooijeweert, Carmen; University Medical Centre Utrecht, Department of Pathology |
| Primary Subject Heading: | Pathology |
| Secondary Subject Heading: | Urology, Diagnostics, Oncology |
| Keywords: | Prostate disease < UROLOGY, Breast tumours < ONCOLOGY, PATHOLOGY |
| | |

# The CONFIDENT-trial protocol: a pragmatic template for clinical implementation of artificial intelligence assistance in pathology


## Authors

R.N. Flach (r.n.flach-2@umcutrecht.nl), N. Stathonikos (nstatho2@umcutrecht.nl), T.Q. Nguyen (t.q.nguyen@umcutrecht.nl), N.D. ter Hoeve (N.terhoeve@umcutrecht.nl), P.J. van Diest (P.j.vandiest@umcutrecht.nl), C. van Dooijeweert (c.vandooijeweert-3@umcutrecht.nl)

## Affiliation

Department of Pathology, University Medical Centre Utrecht, Utrecht, The Netherlands


## Trial registration

Not applicable

## Protocol version

Version 1.


## Funding

No funding was received for this research at the moment of writing this paper. In the meanwhile, the Hanarth Foundation has provided funding to support this study.


## Roles and responsibilities

As no funding was received, and as this is a single centre clinical trial, full authority over all study activities (design; collection, management, analysis and interpretation of data; writing of the report; and the decision to submit the report for publication) lies with the researchers in the UMC Utrecht.


## Corresponding author

Carmen van Dooijeweert

Internal mail no H04.312

T: +31 88 75 565 65

1 c.vandooijeweert3@umcutrecht.nl

2

# Abstract

## Introduction

Artificial Intelligence (AI) has been on the rise in the field of pathology. Despite promising results in retrospective studies, and several CE-IVD certified algorithms on the market, prospective clinical implementation studies of AI have yet to be performed, to the best of our knowledge. In this trial, we will explore the benefits of an AI-assisted pathology workflow, while maintaining diagnostic safety standards.

## Methods and analysis

This is a SPIRIT-AI compliant single-centre, controlled clinical trial, in a fully digital academic pathology laboratory. We will prospectively include prostate cancer patients who undergo prostate needle biopsies (CONFIDENT-P) and breast cancer patients who undergo a sentinel lymph node (SN) procedure (CONFIDENT-B) in the UMC Utrecht. For both the CONFIDENT-B and CONFIDENT-P trials, the specific pathology specimens will be pseudo-randomized to be assessed by a pathologist with- or without AI-assistance in a pragmatic (bi-)weekly sequential design. In the intervention group, pathologists will assess whole slide images (WSI) of the standard haematoxylin-eosin (HE)-stained sections assisted by the output of the algorithm. In the control group, pathologists will assess HE WSI according to the current clinical workflow. If no tumour cells are identified or when the pathologist is in doubt, immunohistochemistry (IHC) staining will be performed. At least 80 patients in the CONFIDENT-P and 180 patients in the CONFIDENT-B trial will need to be enrolled to detect superiority, allocated as 1:1. Primary endpoint for both trials is the number of saved resources of IHC staining procedures for detecting tumour cells, since this will clarify tangible cost savings that will support the business case for AI.

## Ethics and dissemination

The ethics committee (MREC NedMec) waived the need of official ethical approval, since participants are not subjected to procedures nor are they required to follow rules. Results of both trials (CONFIDENT-B and CONFIDENT-P) will be published in scientific peer-reviewed journals.


# Strengths and limitations

- This is the first clinical trial to examine the added value of artificial intelligence in the daily pathology workflow.
- By maintaining the current diagnostic safety standards patients are not at risk of an inferior diagnosis during the trial.
- This is a pragmatic template for prospective AI-trials for object-identifying algorithms in pathology.
- A limitation is that this is a single-centre study, which may hamper generalizability.
- Due to the existing clinical workflow, randomization of patients and (double-)blinding of the participating pathologists and researchers is not possible.

**Keywords:**
Artificial intelligence, prospective trial, prostate cancer, breast cancer, digital pathology

Word count: 2864 words

# Introduction

## Background and rationale

Since the introduction of digital pathology, the number of studies on artificial intelligence (AI) within the field of pathology has increased exponentially.[1,2] Algorithms have been created for tumour detection, tumour grading, tumour subtyping, evaluating biomarkers and more.[1,3] Due to demographic trends, the needs for healthcare are increasing globally which combined with a lack of specialists, increases the current workload .[2,4] Therefore, AI has great potential to alleviate pathologists' workload[2] and improve diagnostics by improving accuracy, reproducibility and speed.[2] In fact, several algorithms have shown to be comparable, or even superior to pathologists (under time-constraint).[2,5–10]

Artificial and human intelligence are not mutually exclusive, they complement each other, a concept which is known as "augmented intelligence", where AI can enhance, rather than replace human intelligence.[11] In the (very) early AI-adoption phase, and presumably also in later phases, pathologist supervision remains of key importance. This is particularly relevant as, despite the promising results of retrospective studies and the availability of CE-IVD approved algorithms, prospective validation and clinical implementation of AI is currently lacking. For example, six years after the successful CAMELYON-16 Grand Challenge[6], the top algorithms have yet to be implemented in daily clinical practice, showing that the time between development of an AI model and clinical implementation is considerable. Likewise, numerous promising prostate cancer grading algorithms have been developed, yet implementation studies have yet to be performed,[12–14] whereas nine AI pathology devices received CE-IVD approval in 2021.[15]

## Trial rationale

As a pathology laboratory with a fully digital workflow for over seven years, we are eager to explore the full potential of working digitally by adding the benefit of AI in daily pathology practice. We decided to start with the object localisation task of tumour-detection, where an objective reference standard is in place in the routine clinical workflow (i.e. pathologist-supervision and/or immunohistochemistry (IHC) staining).

We developed the CONFIDENT-trial template, in which AI tumour detection algorithms can be safely implemented in prospective clinical trials, while ensuring that patients are not at risk of receiving an inferior diagnosis, since IHC is always performed when no tumour cells are visible, but also when pathologists need more confirmation about the diagnosis.

## CONFIDENT-B and CONFIDENT-P

Our trials aim to prospectively investigate the added value of an AI-assisted pathology workflow in the identification of prostate cancer (PCa) in prostate needle biopsies (CONFIDENT-P) and the identification of sentinel lymph node (SN) metastases in breast cancer (BCa) patients (CONFIDENT-B). As both prostate cancer and breast cancer are the most common (non-skin) malignancies in men and women, respectively, implementation of AI-assistance may have a great impact on diagnostic processes.[16] However, it is important to emphasize that this trial serves as a template for other pragmatic AI-intervention trials for object-localisation tasks as well.

We obtained CE-IVD-approved algorithms for detection and grading of prostate cancer in prostate needle biopsies and an algorithm for detecting lymph node metastases in BCa patients. In both cases the task of the pathologist is both labour-intensive and expensive, due to the performed IHC stains in case no tumour cells are morphologically observed. However, IHC is expensive and these costs sometimes even exceed reimbursement for the entire specimen (e.g. in case of multiple blocks of multiple SNs). This raises the question whether AI may be of added value to morphologically detect cancer cells without the need for IHC-use. Thereby the number of performed IHC stains may be reduced, which may lead to tangible costs savings that will help to build the business case for AI, while potentially decreasing the workload of pathologists as well.[2]

### Study objective

The primary objective is to explore whether an AI-assisted workflow reduces the number of spent resources on IHC, while maintaining diagnostic safety standards in both PCa patients who underwent prostate needle biopsies (CONFIDENT-P), and BCa patients who underwent an SN procedure.

Secondary objectives are to investigate whether time management improves in an AI-assisted workflow and to analyse how many IHC staining may have been safely omitted after AI-implementation.

## Methods and analysis

### Trial design

The study protocol is structured following the SPIRIT-AI (Standard Protocol Items: recommendations for Interventional Trials - Artificial Intelligence) statement 2020[17]. This study is a single-centre, parallel-group controlled trial, assessing superiority. The allocation ratio is 1:1. Eligible patients will be assigned to arm 1 (control group) or arm 2 (AI-assisted workflow), based on a bi-weekly time-schedule. Eligibility criteria are summarised in Figure 1. The CONFIDENT trials will be carried out in 2022-2023.

### Study setting

The trial will take place in the daily practice of a single academic hospital (University Medical Centre Utrecht, the Netherlands), with a fully digital pathology enabled clinical set-up, where all slides are digitised using ultrafast whole slide image (WSI) using Hamamatsu S360 scanners and reviewed using the Sectra pathology Picture Archiving and Communication System (PACS). Although the UMC Utrecht is an academic hospital, primary routine pathology diagnostics is performed for non-academic hospitals as well (i.e., the Alexander Monro Breast Cancer Hospital, Bilthoven, the Netherlands).

### Study population

For PCa, WSI of all males who undergo a prostate needle biopsy in the UMC Utrecht will be included. For BCa, WSI of all females or males with BCa as primary malignancy (i.e. invasive breast cancer) who undergo an SN procedure in the Alexander Monro Breast Cancer Hospital or the UMC Utrecht will be included. Patients will be excluded, if they were redirected to the UMC Utrecht for a second opinion.

### Assessment of specimen

During the study period, all WSI will be assessed by the same group of pathologists; i.e. two expert urological pathologists for the PCa biopsies, and three expert breast pathologists for the lymph node assessment from BCa patients.

For both the CONFIDENT-B and CONFIDENT-P trial, the specific pathology specimens will be assigned to be assessed by a pathologist with or without AI-assistance in a pragmatic (bi-)weekly sequential design. This is considered feasible as the change in the case mix and time trends are unlikely to occur within the inclusion-period of about six to nine months. Furthermore, both specialized breast- and urological pathologists within the UMC Utrecht work according to weekly schedules. Therefore, using AI every other week or every other two weeks, as opposed to switching by day, ensures that all pathologists are equally distributed between groups. Lastly, it would be impractical to switch from AI-assistance in the intervention group to no AI-assistance in the control group on a case-to-case basis. For obvious reasons, allocation concealment and blinding of pathologists and researchers is not applicable.

## Control and intervention

All eligible specimens will be assigned to either the control group or the intervention group. In the control group, pathologists will assess H&E stained WSI of patients digitally, according to the current clinical workflow. For PCa biopsies, IHC is routinely performed on all cases. For BCa lymph nodes, if no metastases or tumour are present, IHC staining will be performed. Additional IHC staining will also be performed by additional request of the pathologist in case of doubt.

In the intervention group, pathologists will assess the H&E-specimens digitally with the outcome of the algorithm provided in their first assessment of the specimen. For PCa, they will use the CE-IVD certified Paige Prostate Suite algorithms for tumor detection and tumor volume percentage calculations, which reaches sensitivitiy and specificity of 99% and 93% respectively and which are based on a weakly-supervised deep learning algorithm as described by Campanella et al.[18,19]For BCa, pathologists will use the CE-IVD certified Metastasis Detection App by Visiopharm, a deep-learning algorithm for lymph node metastases of BCa and colon carcinoma with a combined sensitivity and specificity of 98,7 and 99.6% respectively[20]. These algorithms will be integrated within the Sectra PACS where the output of the algorithms will be graphically displayed. AI analysis of the WSI will be performed right after scanning to avoid delays in the clinical workflow. If the AI-assisted pathologist does not detect metastases or tumours on the H&E slide, routine additional IHC staining will be performed by P503S/p63/CK HMW for PCa and CAM5.2 for BCa, to ensure no metastases or tumours are missed. Pathologists can also request an additional IHC if they feel they need this to make an adequate diagnosis (Figure 2).

## Outcome measures

Primary outcome for the CONFIDENT-P trial is the added value of AI-assistance in the detection of PCa and in the detection of tumour volume in prostate needle biopsies in daily pathology practice. The primary outcome for the CONFIDENT-B trial is the added value of AI-assistance in the detection of BCa SN-metastases. The outcome measures for both trials will be the number of spent resources, i.e. the number of IHC-stains performed in both groups.

Secondary outcome measures will be sensitivity and specificity of the AI-assisted pathologist, time spent on WSI analysis, the number of IHC stains that may have been omitted after AI-implementation and a pathologists' evaluation by a questionnaire on the AI-assisted work process. Sensitivity and specificity analyses of the algorithm itself have already been well documented, and is therefore outside the scope of the paper, as we focus on the combination of pathologist and AI to explore cost savings.

## Input data

Input data for the algorithm will be WSI of haematoxylin and eosin (H&E) stained slides scanned at 40x of either prostate needle biopsies, and WSI of H&E stained slides of BCa lymph nodes. As per routine in our daily clinical practice, WSI will be quality controlled after scanning for colour, focus quality and completeness of the scan. When necessary, the specimens are rescanned.

## Sample size

**CONFIDENT-P**

We performed power calculations using a two-sample proportion superiority test, using expected percentages of IHC staining in both study arms. We assume that the pathologists in the control arm can detect 50% of the tumours without using IHC. We expect AI-assisted pathologists to detect 80% of the tumours, without using IHC. These percentages were conservatively derived from the validity study by Raciti (74% for pathologists without AI and 90% for pathologists with AI respectively)[21], by expert pathologist opinion, and taking into account that pathologists under time constraint of daily practice do not detect tumours as well as pathologists without time constraint during retrospective studies[22]. We assume that this effect will be larger for the biopsies assessed without AI than with AI, as AI is assumed to make tumour detection easier.

A sample size of 60 (30 per arm) would give a power of approximately 80%, using a one-sided 5% significance level. However, uncertainties remain regarding the sample size parameters. We therefore inflated our sample size to 80 (40 per arm), in order to ensure study power and allowing us to detect smaller effect sizes.

For detection of tumour volume percentage, we performed a power calculation based on the assumption that AI should be able to replace at least 20% of the IHC stains, in order to be cost-effective. IHC is currently used in 100% of all prostate needle biopsies. Using a power of 80% and a one-sided significance level of 5%, this leads to 27 patients per arm.

**CONFIDENT-B**

Sample size calculations for the CONFIDENT-B trial are based on the assumption that the AI-algorithm can detect all metastases for which currently IHC is used, which are mainly micro-metastases and isolated tumour cells (ITC). Approximately 15% of the SN-specimens in the UMC Utrecht contain a micrometastasis or ITC.

A sample size of 166 patients (83 per arm) with a one-sided 5% significance level therefore results in a power of 80%. Again, as there are uncertainties on the assumptions on what amount of the metastases will be detect by AI, we decided to be conservative and include 180 patients (90 per arm).

Overall, we are only interested in one-sided outcomes, as it is not possible that more IHC will be performed in the AI-assisted arm. IHC is performed to detect metastases, when they are macroscopically undetectable, rather than to confirm them when they are macroscopically visible. As AI would show only more metastases than the pathologist could macroscopically detect, this means that only a reduction of IHC is possible.

## Statistical methods

For baseline comparisons between both arms, the appropriate measures (parametric or non-parametric) for categorical (Chi-squared test/Fisher's exact) and continuous variables (T-test, Mann-

Whitney U test) will be used. For the analysis of the primary outcome measure, we will compare the proportion of IHC-use in both arms, and calculate adjusted relative risks, using a log-binomial model.[23–25]

Missing data for baseline characteristics and for the primary outcome are not to be expected, as they are obligatory items in the structured pathology reports.

We will determine sensitivity and specificity of the conclusion of the AI-assisted pathologists without the use of IHC. Subsequently, we will focus on the cases with metastases that the AI-assisted pathologist misses, categorize them (i.e. macro-metastases, micro-metastases, ITC) and determine their clinical relevance (i.e. clinical consequences if these metastases are being missed). Data-analysis will be performed in R Statistical Software[26], with a significance level set at $p<0.05$.

## Data collection and management

All data (baseline and primary outcome measurements) will be retrieved from the structured pathology reports and will be managed and stored in Castor EDC[27]. For the secondary outcome measure of time spent by the pathologist on a slide, data will be collected on an interval basis for practical reasons, as timing every assessment for months (by stopwatch) was not deemed feasible. For the secondary outcome measure of AI-assisted work process for pathologists, a questionnaire will be distributed to the participating pathologists. The final secondary assessment measurement, the number of IHC stains that may have been omitted after AI-implementation, will be determined by the researchers based on the data from the structured pathology reports (combination of IHC and AI-assisted diagnoses of the pathologist).

## Ethical approval

Research within these trials is not subject to the (Dutch) Medical Research Involving Human Subjects Act (WMO), as participants are not subjected to procedures and as they are not required to follow rules. Therefore, the ethics committee (MREC NedMec) waived the need of ethical approval and informed consent.

## Risk of harm

Patients are not at risk of any harm for an inferior diagnosis (i.e. missed tumour cells), as in both arms, IHC-staining will be performed when no tumour cells are visible, according to current clinical standards. As a rule in augmented intelligence, all cases will be evaluated by a pathologist, which further minimized the risk of a false diagnosis based on the AI algorithm. Taking all of the above into account, a data monitoring committee (DMC) is not required and adverse events are not to be expected. In theory, the algorithm could be more of a disturbance than a help to pathologists (for example, when it frequently reports false positive or negative results, which have to be corrected by the pathologist). However, the algorithms used are IVDR-approved, and thus have undergone extensive review for their intended purpose. Nonetheless, the experience and ease of use of pathologists working with the algorithm will be one of the secondary outcome measures.

## Informed consent and data access

Informed consent was waived by the local quality coordinator (QC) and data protection officer (DPO) for the following reasons. First, in both arms patients receive standard care, while maintaining diagnostic safety standards (pathologists' supervision, IHC in all negative cases). Second, patients are not subjected to any procedures. Third, all patient data will be anonymized to the researchers by the pathologist who assessed the slide.

The collected (anonymous) research data will be stored in Castor EDC to ensure data security. Data will be kept for a period of 15 years. Data access in Castor will be restricted to two researchers (RF, CvD). Pathologists have access to the electronic patient files for the purpose of patient care. The researchers are not permitted access to these files. At no point will the data (both in Castor EDC and patient files) be accessed by the companies providing the algorithms (i.e. Visiopharm and Paige).

## Patient and Public Involvement

None

# Discussion

The promising retrospective results of AI-assisted pathology have not yet resulted in prospective clinical implementation studies. This may be due to a lack of digital transition in the majority of pathology laboratories, but it may also be partly due to the lack of a good implementation model. Fortunately, however, new guidelines for AI-trials have recently been proposed by the SPIRIT-AI and CONSORT-AI-steering groups, as well as roadmaps to routine use of AI in clinical practice.[28–30] Yet, to date, no pathology AI-trials have been published in PubMed or Web of Science or, to the best of our knowledge, otherwise made public.

As a pathology laboratory with a fully digital workflow, we developed a clinical trial template for tumour detection models, as a first step to implement AI in daily pathology practice. We will start with an object localisation task (i.e. tumour cells) as a reference standard is in place in the routine clinical workflow. For classification tasks like tumour grading, a clinical trial design is more challenging, as no reference is in place in daily pathology practice and inter-laboratory and inter-pathologist variation is notorious.[31–35] Nevertheless, in future trials, implementing AI-assistance in the grading process might also reduce this variation. For now, results of the CONFIDENT-trials will provide the first assessment of the potential added value of AI in daily pathology practice. This evaluation will substantially contribute to a potential paradigm shift in tumour detection in pathology. The pragmatic template of the CONFIDENT trials may serve as example for other prospective AI implementation trials in diagnostic pathology.

## Declaration of interests

PJvD is a member of the Scientific Advisory Board of Paige and Sectra.
All other authors do not report conflict of interest.

## Funding

No funding was obtained at the moment of writing this paper. In the meanwhile, the Hanarth Foundation has provided funding to support this study.

## Dissemination policy

Results of both trials (CONFIDENT-B and CONFIDENT-P) will be published in scientific peer-reviewed journals. Authorship will be acknowledged to all those that substantially contributed in the CONFIDENT-trials. Data will be available upon reasonable request.

## Author statement

PJvD conceived of the study. RNF and CvD initiated the study design and NS, TQN and NDtH helped with implementation. RNF and CvD provided statistical expertise in clinical

trial design and RNF and CvD are conducting the primary statistical analysis. All authors contributed to refinement of the study protocol and approved the final manuscript.

## Figure legends

Figure 1. Flowchart with patient selection; SN = sentinel node; AI = artificial intelligence.

Figure 2. Study flow chart. H&E = hematoxylin & eosin; IHC = immunohistochemistry

# References


1 Jiang Y, Yang M, Wang S, *et al.* Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Commun* 2020;**40**:154–66. doi:https://doi.org/10.1002/cac2.12012

2 van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. *Nat Med* 2021;**27**:775–84. doi:10.1038/s41591-021-01343-4

3 Steiner DF, Chen PHC, Mermel CH. Closing the translation gap: AI applications in digital pathology. *Biochim Biophys Acta Rev Cancer* 2021;**1875**:188452. doi:10.1016/j.bbcan.2020.188452

4 Robboy SJ, Weintraub S, Horvath AE, *et al.* Pathologist workforce in the United States: I. Development of a predictive model to examine factors influencing supply. *Arch Pathol Lab Med* 2013;**137**:1723–32. doi:10.5858/arpa.2013-0200-OA

5 Bulten W, Pinckaers H, van Boven H, *et al.* Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol* 2020;**21**:233–41. doi:10.1016/S1470-2045(19)30739-9

6 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al.* Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* 2017;**318**:2199–210. doi:10.1001/jama.2017.14585

7 Steiner DF, Nagpal K, Sayres R, *et al.* Evaluation of the Use of Combined Artificial Intelligence and Pathologist Assessment to Review and Grade Prostate Biopsies. *JAMA Netw Open* 2020;**3**:e2023267. doi:10.1001/jamanetworkopen.2020.23267

8 Ström P, Kartasalo K, Olsson H, *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020;**21**:222–32. doi:10.1016/S1470-2045(19)30738-7

9 Raciti P, Sue J, Ceballos R, *et al.* Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Modern Pathology* 2020;**33**:2058–66. doi:10.1038/s41379-020-0551-y

10 Hekler A, Utikal JS, Enk AH, *et al.* Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer* 2019;**120**:114–21. doi:10.1016/j.ejca.2019.07.019

11 Harrison JH, Gilbertson JR, Hanna MG, *et al.* Introduction to Artificial Intelligence and Machine Learning for Pathology. *Arch Pathol Lab Med* 2021;**145**:1228–54. doi:10.5858/arpa.2020-0541-cp

12 Ström P, Kartasalo K, Olsson H, *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020;**21**:222–32. doi:10.1016/S1470-2045(19)30738-7

13 Steiner DF, Nagpal K, Sayres R, *et al.* Evaluation of the Use of Combined Artificial Intelligence and Pathologist Assessment to Review and Grade Prostate Biopsies. *JAMA Netw Open* 2020;**3**:1–14. doi:10.1001/jamanetworkopen.2020.23267

14 Bulten W, Balkenhol M, Belinga JJA, *et al.* Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. *Modern Pathology* 2021;**34**:660–71. doi:10.1038/s41379-020-0640-y

15 Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health* 2021;**3**:e195–203. doi:10.1016/S2589-7500(20)30292-2

16 Sung H, Ferlay J, Siegel RL, *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;**71**:209–49. doi:10.3322/caac.21660

17 Cruz Rivera S, Liu X, Chan AW, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020;**2**:e549–60. doi:10.1016/S2589-7500(20)30219-3

18 Campanella G, Hanna MG, Geneslaw L, *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019;**25**:1301–9. doi:10.1038/s41591-019-0508-1

19 da Silva LM, Pereira EM, Salles PGO, *et al.* Independent real-world application of a clinical-grade automated prostate cancer detection system. *Journal of Pathology* 2021;**254**:147–58. doi:10.1002/path.5662

20 Visiopharm. https://visiopharm.com/app-center/app/metastasis-detection-ai/.

21 Raciti P, Sue J, Ceballos R, *et al.* Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Modern Pathology* 2020;**33**:2058–66. doi:10.1038/s41379-020-0551-y

22 Bejnordi BE, Veta M, Van Diest PJ, *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA - Journal of the American Medical Association* 2017;**318**:2199–210. doi:10.1001/jama.2017.14585

23 McNutt LA, Wu C, Xue X, *et al.* Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003;**157**:940–3. doi:10.1093/aje/kwg074

24 Lumley T, Kronmal R, Ma S. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. . *UW Biostatistics Working Paper Series 293 [http://www bepress com/uwbiostat/paper293]* 2006.

25 Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am J Epidemiol* 2004;**159**:702–6. doi:10.1093/aje/kwh090

26 R Core Team. R: A Language and Environment for Statistical Computing. 2020.

27 Castor Electronic Data Capture. 2019.

28 Liu X, Cruz Rivera S, Moher D, *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* 2020;**2**:e537–48. doi:10.1016/S2589-7500(20)30218-1

29 Cruz Rivera S, Liu X, Chan AW, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020;**2**:e549–60. doi:10.1016/S2589-7500(20)30219-3

30 Colling R, Pitman H, Oien K, *et al.* Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *Journal of Pathology J Pathol* 2019;**249**:143–50. doi:10.1002/path.5310

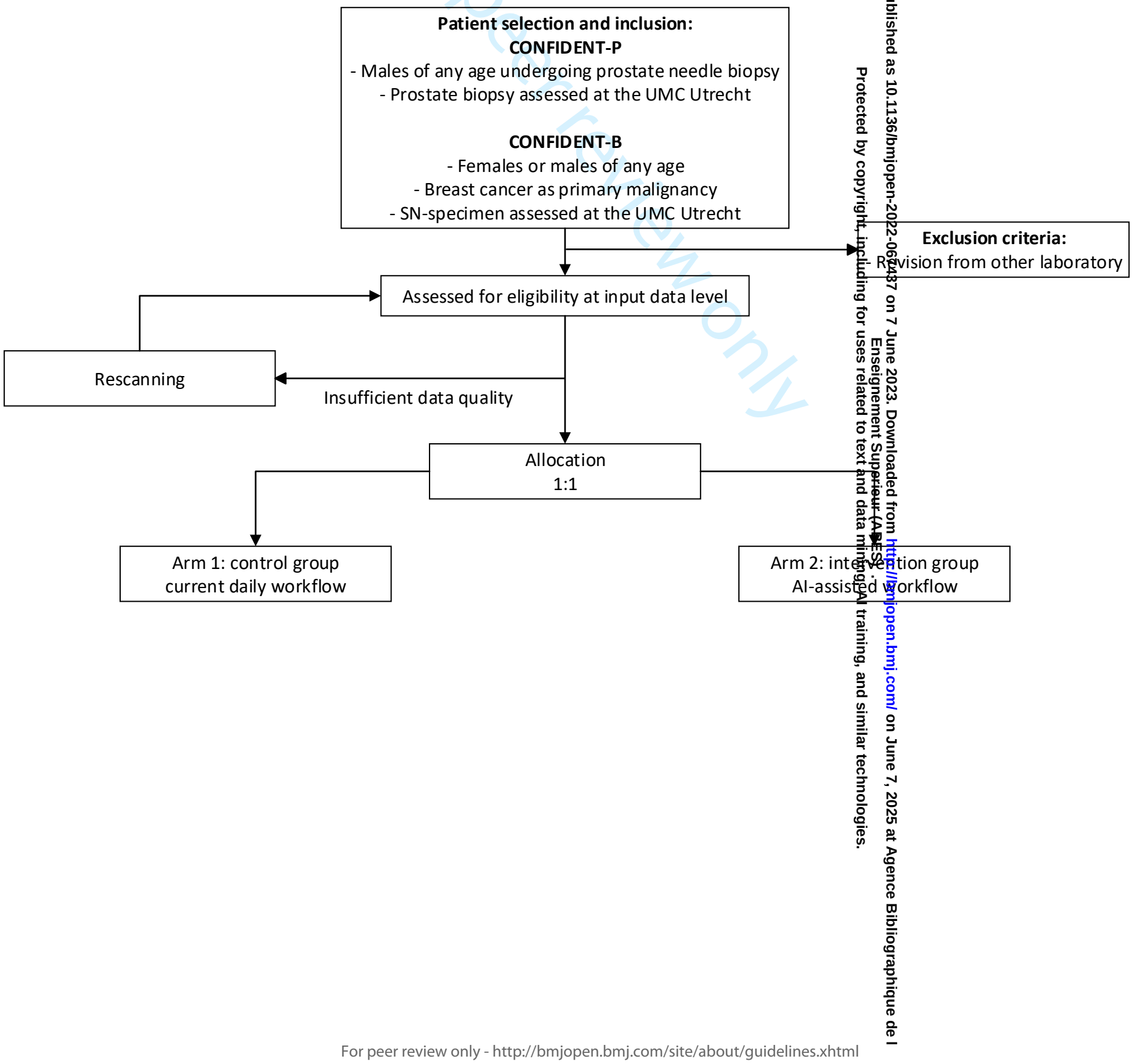
31 Kuijpers CCHJ, Sluijter CE, von der Thüsen JH, *et al.* Interlaboratory variability in the histologic grading of colorectal adenocarcinomas in a nationwide cohort. *American Journal of Surgical Pathology* 2016;**40**:1100–8. doi:10.1111/his.12923

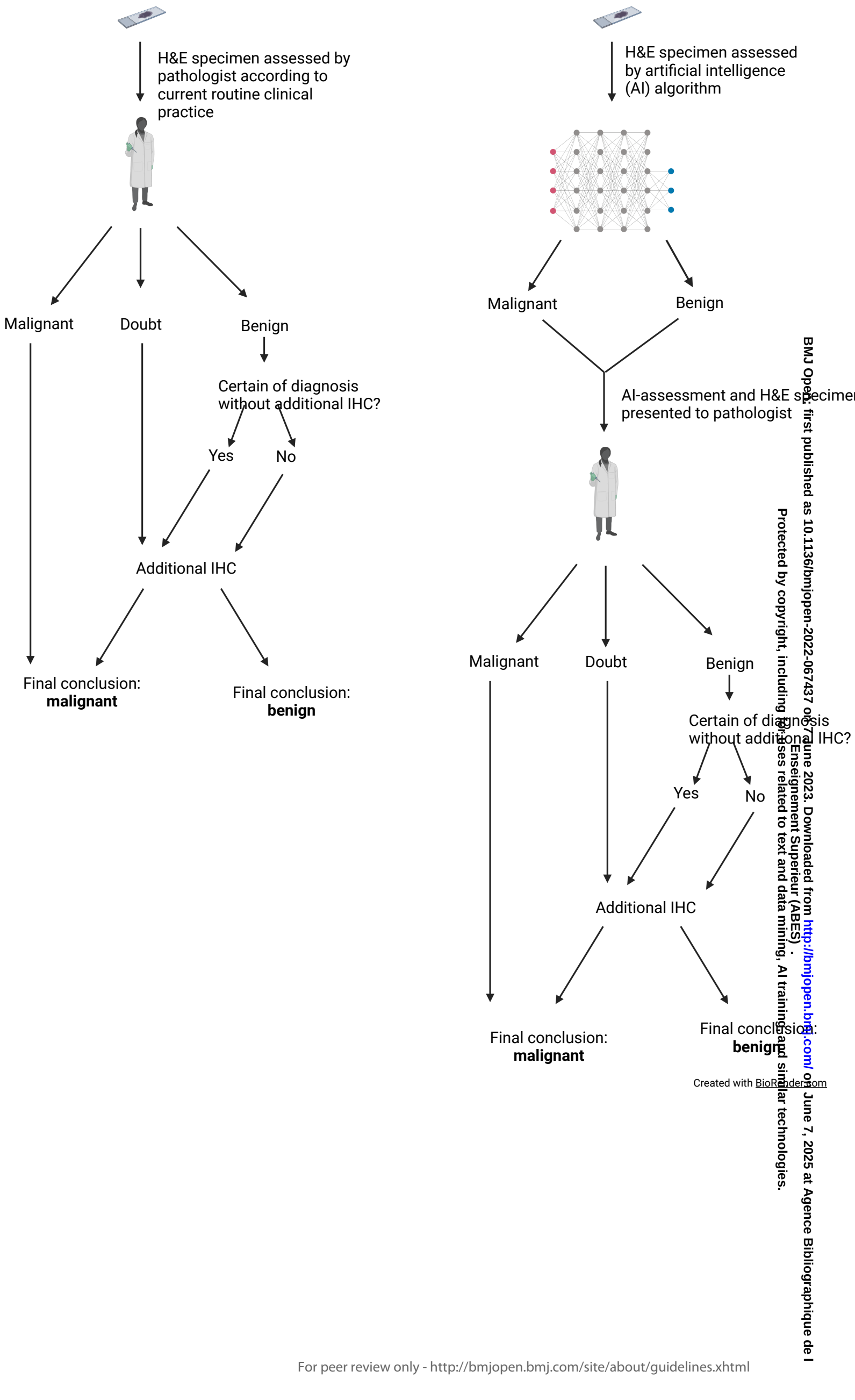
32 Kuijpers CCHJ, Sluijter CE, von der Thüsen JH, *et al.* Interlaboratory variability in the grading of dysplasia in a nationwide cohort of colorectal adenomas. *Histopathology* 2016;**69**:187–97. doi:10.1111/his.12923

33 Flach RN, Willemse P-PM, Suelmann BBM, *et al.* Significant Inter- and Intralaboratory Variation in Gleason Grading of Prostate Cancer: A Nationwide Study of 35,258 Patients in The Netherlands. *Cancers (Basel)* 2021;**13**. doi:10.3390/cancers13215378

34 van Dooijeweert C, van Diest PJ, Willems SM, *et al.* Significant inter- and intra-laboratory variation in grading of ductal carcinoma in situ of the breast: a nationwide study of 4901 patients in the Netherlands. *Breast Cancer Res Treat* 2019;**174**:479–88. doi:10.1007/s10549-018-05082-y

35 van Dooijeweert C, van Diest PJ, Willems SM, *et al.* Significant inter- and intra-laboratory variation in grading of invasive breast cancer: A nationwide study of 33,043 patients in the Netherlands. *Int J Cancer* 2020;**146**:769–80. doi:10.1002/ijc.32330





# Reporting checklist for protocol of a clinical trial.

Based on the SPIRIT guidelines.

## Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the SPIRITreporting guidelines, and cite them as:

Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin J, Dickersin K, Hróbjartsson A, Schulz KF, Parulekar WR, Krleža-Jerić K, Laupacis A, Moher D. SPIRIT 2013 Explanation and Elaboration: Guidance for protocols of clinical trials. BMJ. 2013;346:e7586

| | | Reporting Item | Page Number |
|---|---|---|---|
| **Administrative information** | | | |
| Title | #1 | Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym | 1 |
| Trial registration | #2a | Trial identifier and registry name. If not yet registered, name of intended registry | na |
| Trial registration: data set | #2b | All items from the World Health Organization Trial Registration Data Set | na |
| Protocol version | #3 | Date and version identifier | 1 |
| Funding | #4 | Sources and types of financial, material, and other support | 1 |
| Roles and responsibilities: contributorship | #5a | Names, affiliations, and roles of protocol contributors | 1 |

| | | | |
|---|---|---|---|
| Roles and responsibilities: sponsor contact information | #5b | Name and contact information for the trial sponsor | na |
| Roles and responsibilities: sponsor and funder | #5c | Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities | na |
| Roles and responsibilities: committees | #5d | Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) | na |
| **Introduction** | | | |
| Background and rationale | #6a | Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention | 3 |
| Background and rationale: choice of comparators | #6b | Explanation for choice of comparators | 3 |
| Objectives | #7 | Specific objectives or hypotheses | 4 |
| Trial design | #8 | Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, non-inferiority, exploratory) | 4 |
| **Methods: Participants, interventions, and outcomes** | | | |
| Study setting | #9 | Description of study settings (eg, community clinic, academic hospital) and list of countries where data will | 4-5 |

| | | | |
|---|---|---|---|
| | | be collected. Reference to where list of study sites can be obtained | |
| Eligibility criteria | #10 | Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) | 5 |
| Interventions: description | #11a | Interventions for each group with sufficient detail to allow replication, including how and when they will be administered | 6 |
| Interventions: modifications | #11b | Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving / worsening disease) | na |
| Interventions: adherance | #11c | Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return; laboratory tests) | na |
| Interventions: concomitant care | #11d | Relevant concomitant care and interventions that are permitted or prohibited during the trial | na |
| Outcomes | #12 | Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended | 6, 8 |
| Participant timeline | #13 | Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure) | 6 |
| Sample size | #14 | Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations | 7 |
| Recruitment | #15 | Strategies for achieving adequate participant enrolment to reach target sample size | 5 |

| | | | |
|---|---|---|---|
| **Methods: Assignment of interventions (for controlled trials)** | | | |
| Allocation: sequence generation | #16a | Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions | 5 |
| Allocation concealment mechanism | #16b | Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned | na |
| Allocation: implementation | #16c | Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions | na |
| Blinding (masking) | #17a | Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how | na |
| Blinding (masking): emergency unblinding | #17b | If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial | na |
| **Methods: Data collection, management, and analysis** | | | |
| Data collection plan | #18a | Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. | 8 |

| | | | |
|---|---|---|---|
| | | Reference to where data collection forms can be found, if not in the protocol | |
| Data collection plan: retention | #18b | Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols | na |
| Data management | #19 | Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol | 8-9 |
| Statistics: outcomes | #20a | Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol | 7-8 |
| Statistics: additional analyses | #20b | Methods for any additional analyses (eg, subgroup and adjusted analyses) | 7-8 |
| Statistics: analysis population and missing data | #20c | Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) | 7-8 |
| **Methods: Monitoring** | | | |
| Data monitoring: formal committee | #21a | Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed | 8 |
| Data monitoring: interim analysis | #21b | Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial | na |
| Harms | #22 | Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events | 8 |

| | | | |
|---|---|---|---|
| | | and other unintended effects of trial interventions or trial conduct | |
| Auditing | #23 | Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor | na |
| **Ethics and dissemination** | | | |
| Research ethics approval | #24 | Plans for seeking research ethics committee / institutional review board (REC / IRB) approval | 9 |
| Protocol amendments | #25 | Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC / IRBs, trial participants, trial registries, journals, regulators) | na |
| Consent or assent | #26a | Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) | na |
| Consent or assent: ancillary studies | #26b | Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable | 8 |
| Confidentiality | #27 | How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial | 8 |
| Declaration of interests | #28 | Financial and other competing interests for principal investigators for the overall trial and each study site | 9 |
| Data access | #29 | Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators | 8-9 |
| Ancillary and post trial care | #30 | Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation | na |
| Dissemination policy: trial results | #31a | Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the | 9 |

| | | | |
|---|---|---|---|
| | | public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions | |
| Dissemination policy: authorship | #31b | Authorship eligibility guidelines and any intended use of professional writers | 9 |
| Dissemination policy: reproducible research | #31c | Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code | 9 |
| **Appendices** | | | |
| Informed consent materials | #32 | Model consent form and other related documentation given to participants and authorised surrogates | 8 |
| Biological specimens | #33 | Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable | na |

BMJ Open

# The CONFIDENT-trial protocol: a pragmatic template for clinical implementation of artificial intelligence assistance in pathology



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2022-067437.R2 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 28-Apr-2023 |
| Complete List of Authors: | Flach, Rachel; University Medical Centre Utrecht, Department of Pathology<br>Stathonikos, Nikolas; University Medical Centre Utrecht, Department of Pathology<br>Nguyen, Tri; University Medical Centre Utrecht, Department of Pathology<br>Ter Hoeve, Natalie; University Medical Centre Utrecht, Department of Pathology<br>van Diest, P.J.; University Medical Centre Utrecht, Department of Pathology<br>van Dooijeweert, Carmen; University Medical Centre Utrecht, Department of Pathology |
| Primary Subject Heading: | Pathology |
| Secondary Subject Heading: | Urology, Diagnostics, Oncology |
| Keywords: | Prostate disease < UROLOGY, Breast tumours < ONCOLOGY, PATHOLOGY |
| | |

# The CONFIDENT-trial protocol: a pragmatic template for clinical implementation of artificial intelligence assistance in pathology


## Authors

R.N. Flach (r.n.flach-2@umcutrecht.nl), N. Stathonikos (nstatho2@umcutrecht.nl), T.Q. Nguyen (t.q.nguyen@umcutrecht.nl), N.D. ter Hoeve (N.terhoeve@umcutrecht.nl), P.J. van Diest (P.j.vandiest@umcutrecht.nl), C. van Dooijeweert (c.vandooijeweert-3@umcutrecht.nl)

## Affiliation

Department of Pathology, University Medical Centre Utrecht, Utrecht, The Netherlands


## Trial registration

Not applicable

## Protocol version

Version 1.


## Funding

No funding was received for this research at the moment of writing this paper. In the meanwhile, the Hanarth Foundation has provided funding to support this study.


## Roles and responsibilities

As no funding was received, and as this is a single centre clinical trial, full authority over all study activities (design; collection, management, analysis and interpretation of data; writing of the report; and the decision to submit the report for publication) lies with the researchers in the UMC Utrecht.


## Corresponding author

Carmen van Dooijeweert

Internal mail no H04.312

T: +31 88 75 565 65

1 c.vandooijeweert3@umcutrecht.nl

2

# Abstract

## Introduction

Artificial Intelligence (AI) has been on the rise in the field of pathology. Despite promising results in retrospective studies, and several CE-IVD certified algorithms on the market, prospective clinical implementation studies of AI have yet to be performed, to the best of our knowledge. In this trial, we will explore the benefits of an AI-assisted pathology workflow, while maintaining diagnostic safety standards.

## Methods and analysis

This is a SPIRIT-AI compliant single-centre, controlled clinical trial, in a fully digital academic pathology laboratory. We will prospectively include prostate cancer patients who undergo prostate needle biopsies (CONFIDENT-P) and breast cancer patients who undergo a sentinel lymph node (SN) procedure (CONFIDENT-B) in the UMC Utrecht. For both the CONFIDENT-B and CONFIDENT-P trials, the specific pathology specimens will be pseudo-randomized to be assessed by a pathologist with- or without AI-assistance in a pragmatic (bi-)weekly sequential design. In the intervention group, pathologists will assess whole slide images (WSI) of the standard haematoxylin-eosin (HE)-stained sections assisted by the output of the algorithm. In the control group, pathologists will assess HE WSI according to the current clinical workflow. If no tumour cells are identified or when the pathologist is in doubt, immunohistochemistry (IHC) staining will be performed. At least 80 patients in the CONFIDENT-P and 180 patients in the CONFIDENT-B trial will need to be enrolled to detect superiority, allocated as 1:1. Primary endpoint for both trials is the number of saved resources of IHC staining procedures for detecting tumour cells, since this will clarify tangible cost savings that will support the business case for AI.

## Ethics and dissemination

The ethics committee (MREC NedMec) waived the need of official ethical approval, since participants are not subjected to procedures nor are they required to follow rules. Results of both trials (CONFIDENT-B and CONFIDENT-P) will be published in scientific peer-reviewed journals.


# Strengths and limitations

- This is the first clinical trial to examine the added value of artificial intelligence in the daily pathology workflow.
- By maintaining the current diagnostic safety standards patients are not at risk of an inferior diagnosis during the trial.
- This is a pragmatic template for prospective AI-trials for object-identifying algorithms in pathology.
- A limitation is that this is a single-centre study, which may hamper generalizability.
- Due to the existing clinical workflow, randomization of patients and (double-)blinding of the participating pathologists and researchers is not possible.

**Keywords:**
Artificial intelligence, prospective trial, prostate cancer, breast cancer, digital pathology

Word count: 2864 words

# Introduction

## Background and rationale

Since the introduction of digital pathology, the number of studies on artificial intelligence (AI) within the field of pathology has increased exponentially.[1,2] Algorithms have been created for tumour detection, tumour grading, tumour subtyping, evaluating biomarkers and more.[1,3] Due to demographic trends, the needs for healthcare are increasing globally which combined with a lack of specialists, increases the current workload .[2,4] Therefore, AI has great potential to alleviate pathologists' workload[2] and improve diagnostics by improving accuracy, reproducibility and speed.[2] In fact, several algorithms have shown to be comparable, or even superior to pathologists (under time-constraint).[2,5–10]

Artificial and human intelligence are not mutually exclusive, they complement each other, a concept which is known as "augmented intelligence", where AI can enhance, rather than replace human intelligence.[11] In the (very) early AI-adoption phase, and presumably also in later phases, pathologist supervision remains of key importance. This is particularly relevant as, despite the promising results of retrospective studies and the availability of CE-IVD approved algorithms, prospective validation and clinical implementation of AI is currently lacking. For example, six years after the successful CAMELYON-16 Grand Challenge[6], the top algorithms have yet to be implemented in daily clinical practice, showing that the time between development of an AI model and clinical implementation is considerable. Likewise, numerous promising prostate cancer grading algorithms have been developed, yet implementation studies have yet to be performed,[12–14] whereas nine AI pathology devices received CE-IVD approval in 2021.[15]

## Trial rationale

As a pathology laboratory with a fully digital workflow for over seven years, we are eager to explore the full potential of working digitally by adding the benefit of AI in daily pathology practice. We decided to start with the object localisation task of tumour-detection, where an objective reference standard is in place in the routine clinical workflow (i.e. pathologist-supervision and/or immunohistochemistry (IHC) staining).

We developed the CONFIDENT-trial template, in which AI tumour detection algorithms can be safely implemented in prospective clinical trials, while ensuring that patients are not at risk of receiving an inferior diagnosis, since IHC is always performed when no tumour cells are visible, but also when pathologists need more confirmation about the diagnosis.

## CONFIDENT-B and CONFIDENT-P

Our trials aim to prospectively investigate the added value of an AI-assisted pathology workflow in the identification of prostate cancer (PCa) in prostate needle biopsies (CONFIDENT-P) and the identification of sentinel lymph node (SN) metastases in breast cancer (BCa) patients (CONFIDENT-B). As both prostate cancer and breast cancer are the most common (non-skin) malignancies in men and women, respectively, implementation of AI-assistance may have a great impact on diagnostic processes.[16] However, it is important to emphasize that this trial serves as a template for other pragmatic AI-intervention trials for object-localisation tasks as well.

We obtained CE-IVD-approved algorithms for detection and grading of prostate cancer in prostate needle biopsies and an algorithm for detecting lymph node metastases in BCa patients. In both cases the task of the pathologist is both labour-intensive and expensive, due to the performed IHC stains in case no tumour cells are morphologically observed. However, IHC is expensive and these costs sometimes even exceed reimbursement for the entire specimen (e.g. in case of multiple blocks of multiple SNs). This raises the question whether AI may be of added value to morphologically detect cancer cells without the need for IHC-use. Thereby the number of performed IHC stains may be reduced, which may lead to tangible costs savings that will help to build the business case for AI, while potentially decreasing the workload of pathologists as well.[2]

### Study objective

The primary objective is to explore whether an AI-assisted workflow reduces the number of spent resources on IHC, while maintaining diagnostic safety standards in both PCa patients who underwent prostate needle biopsies (CONFIDENT-P), and BCa patients who underwent an SN procedure.

Secondary objectives are to investigate whether time management improves in an AI-assisted workflow and to analyse how many IHC staining may have been safely omitted after AI-implementation.

## Methods and analysis

### Trial design

The study protocol is structured following the SPIRIT-AI (Standard Protocol Items: recommendations for Interventional Trials - Artificial Intelligence) statement 2020[17]. This study is a single-centre, parallel-group controlled trial, assessing superiority. The allocation ratio is 1:1. Eligible patients will be assigned to arm 1 (control group) or arm 2 (AI-assisted workflow), based on a bi-weekly time-schedule. Eligibility criteria are summarised in Figure 1. The CONFIDENT trials will be carried out in 2022-2023.

### Study setting

The trial will take place in the daily practice of a single academic hospital (University Medical Centre Utrecht, the Netherlands), with a fully digital pathology enabled clinical set-up, where all slides are digitised using ultrafast whole slide image (WSI) using Hamamatsu S360 scanners and reviewed using the Sectra pathology Picture Archiving and Communication System (PACS). Although the UMC Utrecht is an academic hospital, primary routine pathology diagnostics is performed for non-academic hospitals as well (i.e., the Alexander Monro Breast Cancer Hospital, Bilthoven, the Netherlands).

### Study population

For PCa, WSI of all males who undergo a prostate needle biopsy in the UMC Utrecht will be included. For BCa, WSI of all females or males with BCa as primary malignancy (i.e. invasive breast cancer) who undergo an SN procedure in the Alexander Monro Breast Cancer Hospital or the UMC Utrecht will be included. Patients will be excluded, if they were redirected to the UMC Utrecht for a second opinion.

### Assessment of specimen

During the study period, all WSI will be assessed by the same group of pathologists; i.e. two expert urological pathologists for the PCa biopsies, and three expert breast pathologists for the lymph node assessment from BCa patients.

For both the CONFIDENT-B and CONFIDENT-P trial, the specific pathology specimens will be assigned to be assessed by a pathologist with or without AI-assistance in a pragmatic (bi-)weekly sequential design. This is considered feasible as the change in the case mix and time trends are unlikely to occur within the inclusion-period of about six to nine months. Furthermore, both specialized breast- and urological pathologists within the UMC Utrecht work according to weekly schedules. Therefore, using AI every other week or every other two weeks, as opposed to switching by day, ensures that all pathologists are equally distributed between groups. Lastly, it would be impractical to switch from AI-assistance in the intervention group to no AI-assistance in the control group on a case-to-case basis. For obvious reasons, allocation concealment and blinding of pathologists and researchers is not applicable.

## Control and intervention

All eligible specimens will be assigned to either the control group or the intervention group. In the control group, pathologists will assess H&E stained WSI of patients digitally, according to the current clinical workflow. For PCa biopsies, IHC is routinely performed on all cases. For BCa lymph nodes, if no metastases or tumour are present, IHC staining will be performed. Additional IHC staining will also be performed by additional request of the pathologist in case of doubt.

In the intervention group, pathologists will assess the H&E-specimens digitally with the outcome of the algorithm provided in their first assessment of the specimen. For PCa, they will use the CE-IVD certified Paige Prostate Suite algorithms for tumor detection and tumor volume percentage calculations, which reaches sensitivitiy and specificity of 99% and 93% respectively and which are based on a weakly-supervised deep learning algorithm as described by Campanella et al.[18,19]For BCa, pathologists will use the CE-IVD certified Metastasis Detection App by Visiopharm, a deep-learning algorithm for lymph node metastases of BCa and colon carcinoma with a combined sensitivity and specificity of 98,7 and 99.6% respectively[20]. These algorithms will be integrated within the Sectra PACS where the output of the algorithms will be graphically displayed. AI analysis of the WSI will be performed right after scanning to avoid delays in the clinical workflow. If the AI-assisted pathologist does not detect metastases or tumours on the H&E slide, routine additional IHC staining will be performed by P503S/p63/CK HMW for PCa and CAM5.2 for BCa, to ensure no metastases or tumours are missed. Pathologists can also request an additional IHC if they feel they need this to make an adequate diagnosis (Figure 2).

## Outcome measures

Primary outcome for the CONFIDENT-P trial is the added value of AI-assistance in the detection of PCa and in the detection of tumour volume in prostate needle biopsies in daily pathology practice. The primary outcome for the CONFIDENT-B trial is the added value of AI-assistance in the detection of BCa SN-metastases. The outcome measures for both trials will be the number of spent resources, i.e. the number of IHC-stains performed in both groups.

Secondary outcome measures will be sensitivity and specificity of the AI-assisted pathologist, time spent on WSI analysis, the number of IHC stains that may have been omitted after AI-implementation and a pathologists' evaluation by a questionnaire on the AI-assisted work process. Sensitivity and specificity analyses of the algorithm itself have already been well documented, and is therefore outside the scope of the paper, as we focus on the combination of pathologist and AI to explore cost savings.

## Input data

Input data for the algorithm will be WSI of haematoxylin and eosin (H&E) stained slides scanned at 40x of either prostate needle biopsies, and WSI of H&E stained slides of BCa lymph nodes. As per routine in our daily clinical practice, WSI will be quality controlled after scanning for colour, focus quality and completeness of the scan. When necessary, the specimens are rescanned.

## Sample size

**CONFIDENT-P**

We performed power calculations using a two-sample proportion superiority test, using expected percentages of IHC staining in both study arms. We assume that the pathologists in the control arm can detect 50% of the tumours without using IHC. We expect AI-assisted pathologists to detect 80% of the tumours, without using IHC. These percentages were conservatively derived from the validity study by Raciti (74% for pathologists without AI and 90% for pathologists with AI respectively)[21], by expert pathologist opinion, and taking into account that pathologists under time constraint of daily practice do not detect tumours as well as pathologists without time constraint during retrospective studies[22]. We assume that this effect will be larger for the biopsies assessed without AI than with AI, as AI is assumed to make tumour detection easier.

A sample size of 60 (30 per arm) would give a power of approximately 80%, using a one-sided 5% significance level. However, uncertainties remain regarding the sample size parameters. We therefore inflated our sample size to 80 (40 per arm), in order to ensure study power and allowing us to detect smaller effect sizes.

For detection of tumour volume percentage, we performed a power calculation based on the assumption that AI should be able to replace at least 20% of the IHC stains, in order to be cost-effective. IHC is currently used in 100% of all prostate needle biopsies. Using a power of 80% and a one-sided significance level of 5%, this leads to 27 patients per arm.

**CONFIDENT-B**

Sample size calculations for the CONFIDENT-B trial are based on the assumption that the AI-algorithm can detect all metastases for which currently IHC is used, which are mainly micro-metastases and isolated tumour cells (ITC). Approximately 15% of the SN-specimens in the UMC Utrecht contain a micrometastasis or ITC.

A sample size of 166 patients (83 per arm) with a one-sided 5% significance level therefore results in a power of 80%. Again, as there are uncertainties on the assumptions on what amount of the metastases will be detect by AI, we decided to be conservative and include 180 patients (90 per arm).

Overall, we are only interested in one-sided outcomes, as it is not possible that more IHC will be performed in the AI-assisted arm. IHC is performed to detect metastases, when they are macroscopically undetectable, rather than to confirm them when they are macroscopically visible. As AI would show only more metastases than the pathologist could macroscopically detect, this means that only a reduction of IHC is possible.

Sample sizes were calculated uwing the power.prop.test command in R version 4.2.2[23].

## Statistical methods

For baseline comparisons between both arms, the appropriate measures (parametric or non-parametric) for categorical (Chi-squared test/Fisher's exact) and continuous variables (T-test, Mann-Whitney U test) will be used. For the analysis of the primary outcome measure, we will compare the proportion of IHC-use in both arms, and calculate adjusted relative risks, using a log-binomial model.[24–26]

Missing data for baseline characteristics and for the primary outcome are not to be expected, as they are obligatory items in the structured pathology reports.

We will determine sensitivity and specificity of the conclusion of the AI-assisted pathologists without the use of IHC. Subsequently, we will focus on the cases with metastases that the AI-assisted pathologist misses, categorize them (i.e. macro-metastases, micro-metastases, ITC) and determine their clinical relevance (i.e. clinical consequences if these metastases are being missed). Data-analysis will be performed in R Statistical Software[23], with a significance level set at $p<0.05$.

## Data collection and management

All data (baseline and primary outcome measurements) will be retrieved from the structured pathology reports and will be managed and stored in Castor EDC[27]. For the secondary outcome measure of time spent by the pathologist on a slide, data will be collected on an interval basis for practical reasons, as timing every assessment for months (by stopwatch) was not deemed feasible. For the secondary outcome measure of AI-assisted work process for pathologists, a questionnaire will be distributed to the participating pathologists. The final secondary assessment measurement, the number of IHC stains that may have been omitted after AI-implementation, will be determined by the researchers based on the data from the structured pathology reports (combination of IHC and AI-assisted diagnoses of the pathologist).

## Ethical approval

Research within these trials is not subject to the (Dutch) Medical Research Involving Human Subjects Act (WMO), as participants are not subjected to procedures and as they are not required to follow rules. Therefore, the ethics committee (MREC NedMec) waived the need of ethical approval and informed consent.

## Risk of harm

Patients are not at risk of any harm for an inferior diagnosis (i.e. missed tumour cells), as in both arms, IHC-staining will be performed when no tumour cells are visible, according to current clinical standards. As a rule in augmented intelligence, all cases will be evaluated by a pathologist, which further minimized the risk of a false diagnosis based on the AI algorithm. Taking all of the above into account, a data monitoring committee (DMC) is not required and adverse events are not to be expected. In theory, the algorithm could be more of a disturbance than a help to pathologists (for example, when it frequently reports false positive or negative results, which have to be corrected by the pathologist). However, the algorithms used are IVDR-approved, and thus have undergone extensive review for their intended purpose. Nonetheless, the experience and ease of use of pathologists working with the algorithm will be one of the secondary outcome measures.

## Informed consent and data access

Informed consent was waived by the local quality coordinator (QC) and data protection officer (DPO) for the following reasons. First, in both arms patients receive standard care, while maintaining diagnostic safety standards (pathologists' supervision, IHC in all negative cases). Second, patients are not subjected to any procedures. Third, all patient data will be anonymized to the researchers by the pathologist who assessed the slide.

The collected (anonymous) research data will be stored in Castor EDC to ensure data security. Data will be kept for a period of 15 years. Data access in Castor will be restricted to two researchers (RF, CvD). Pathologists have access to the electronic patient files for the purpose of patient care. The researchers are not permitted access to these files. At no point will the data (both in Castor EDC and patient files) be accessed by the companies providing the algorithms (i.e. Visiopharm and Paige).

## Patient and Public Involvement

None

# Discussion

The promising retrospective results of AI-assisted pathology have not yet resulted in prospective clinical implementation studies. This may be due to a lack of digital transition in the majority of pathology laboratories, but it may also be partly due to the lack of a good implementation model. Fortunately, however, new guidelines for AI-trials have recently been proposed by the SPIRIT-AI and CONSORT-AI-steering groups, as well as roadmaps to routine use of AI in clinical practice.[28–30] Yet, to date, no pathology AI-trials have been published in PubMed or Web of Science or, to the best of our knowledge, otherwise made public.

As a pathology laboratory with a fully digital workflow, we developed a clinical trial template for tumour detection models, as a first step to implement AI in daily pathology practice. We will start with an object localisation task (i.e. tumour cells) as a reference standard is in place in the routine clinical workflow. For classification tasks like tumour grading, a clinical trial design is more challenging, as no reference is in place in daily pathology practice and inter-laboratory and inter-pathologist variation is notorious.[31–35] Nevertheless, in future trials, implementing AI-assistance in the grading process might also reduce this variation. For now, results of the CONFIDENT-trials will provide the first assessment of the potential added value of AI in daily pathology practice. This evaluation will substantially contribute to a potential paradigm shift in tumour detection in pathology. The pragmatic template of the CONFIDENT trials may serve as example for other prospective AI implementation trials in diagnostic pathology.

## Declaration of interests

PJvD is a member of the Scientific Advisory Board of Paige and Sectra.
All other authors do not report conflict of interest.

## Funding

No funding was obtained at the moment of writing this paper. In the meanwhile, the Hanarth Foundation has provided funding to support this study.

## Dissemination policy

Results of both trials (CONFIDENT-B and CONFIDENT-P) will be published in scientific peer-reviewed journals. Authorship will be acknowledged to all those that substantially contributed in the CONFIDENT-trials. Data will be available upon reasonable request.

## Author statement

PJvD conceived of the study. RNF and CvD initiated the study design and NS, TQN and NDtH helped with implementation. RNF and CvD provided statistical expertise in clinical trial design and RNF and CvD are conducting the primary statistical analysis. All authors contributed to refinement of the study protocol and approved the final manuscript.

## Figure legends

Figure 1. Flowchart with patient selection; SN = sentinel node; AI = artificial intelligence.

Figure 2. Study flow chart. H&E = hematoxylin & eosin; IHC = immunohistochemistry

# References


1 Jiang Y, Yang M, Wang S, *et al.* Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Commun* 2020;**40**:154–66. doi:https://doi.org/10.1002/cac2.12012

2 van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. *Nat Med* 2021;**27**:775–84. doi:10.1038/s41591-021-01343-4

3 Steiner DF, Chen PHC, Mermel CH. Closing the translation gap: AI applications in digital pathology. *Biochim Biophys Acta Rev Cancer* 2021;**1875**:188452. doi:10.1016/j.bbcan.2020.188452

4 Robboy SJ, Weintraub S, Horvath AE, *et al.* Pathologist workforce in the United States: I. Development of a predictive model to examine factors influencing supply. *Arch Pathol Lab Med* 2013;**137**:1723–32. doi:10.5858/arpa.2013-0200-OA

5 Bulten W, Pinckaers H, van Boven H, *et al.* Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol* 2020;**21**:233–41. doi:10.1016/S1470-2045(19)30739-9

6 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al.* Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* 2017;**318**:2199–210. doi:10.1001/jama.2017.14585

7 Steiner DF, Nagpal K, Sayres R, *et al.* Evaluation of the Use of Combined Artificial Intelligence and Pathologist Assessment to Review and Grade Prostate Biopsies. *JAMA Netw Open* 2020;**3**:e2023267. doi:10.1001/jamanetworkopen.2020.23267

8 Ström P, Kartasalo K, Olsson H, *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020;**21**:222–32. doi:10.1016/S1470-2045(19)30738-7

9 Raciti P, Sue J, Ceballos R, *et al.* Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Modern Pathology* 2020;**33**:2058–66. doi:10.1038/s41379-020-0551-y

10 Hekler A, Utikal JS, Enk AH, *et al.* Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer* 2019;**120**:114–21. doi:10.1016/j.ejca.2019.07.019

11 Harrison JH, Gilbertson JR, Hanna MG, *et al.* Introduction to Artificial Intelligence and Machine Learning for Pathology. *Arch Pathol Lab Med* 2021;**145**:1228–54. doi:10.5858/arpa.2020-0541-cp

12 Ström P, Kartasalo K, Olsson H, *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020;**21**:222–32. doi:10.1016/S1470-2045(19)30738-7

13 Steiner DF, Nagpal K, Sayres R, *et al.* Evaluation of the Use of Combined Artificial Intelligence and Pathologist Assessment to Review and Grade Prostate Biopsies. *JAMA Netw Open* 2020;**3**:1–14. doi:10.1001/jamanetworkopen.2020.23267

14 Bulten W, Balkenhol M, Belinga JJA, *et al.* Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. *Modern Pathology* 2021;**34**:660–71. doi:10.1038/s41379-020-0640-y

15 Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health* 2021;**3**:e195–203. doi:10.1016/S2589-7500(20)30292-2

16 Sung H, Ferlay J, Siegel RL, *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;**71**:209–49. doi:10.3322/caac.21660

17 Cruz Rivera S, Liu X, Chan AW, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020;**2**:e549–60. doi:10.1016/S2589-7500(20)30219-3

18 Campanella G, Hanna MG, Geneslaw L, *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019;**25**:1301–9. doi:10.1038/s41591-019-0508-1

19 da Silva LM, Pereira EM, Salles PGO, *et al.* Independent real-world application of a clinical-grade automated prostate cancer detection system. *Journal of Pathology* 2021;**254**:147–58. doi:10.1002/path.5662

20 Visiopharm. https://visiopharm.com/app-center/app/metastasis-detection-ai/.

21 Raciti P, Sue J, Ceballos R, *et al.* Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Modern Pathology* 2020;**33**:2058–66. doi:10.1038/s41379-020-0551-y

22 Bejnordi BE, Veta M, Van Diest PJ, *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA - Journal of the American Medical Association* 2017;**318**:2199–210. doi:10.1001/jama.2017.14585

23 R Core Team. R: A Language and Environment for Statistical Computing. 2022.https://www.r-project.org/

24 McNutt LA, Wu C, Xue X, *et al.* Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003;**157**:940–3. doi:10.1093/aje/kwg074

25 Lumley T, Kronmal R, Ma S. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. . *UW Biostatistics Working Paper Series 293 [http://www bepress com/uwbiostat/paper293]* 2006.

26 Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am J Epidemiol* 2004;**159**:702–6. doi:10.1093/aje/kwh090

27 Castor Electronic Data Capture. 2019.

28 Liu X, Cruz Rivera S, Moher D, *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* 2020;**2**:e537–48. doi:10.1016/S2589-7500(20)30218-1

29 Cruz Rivera S, Liu X, Chan AW, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020;**2**:e549–60. doi:10.1016/S2589-7500(20)30219-3

30 Colling R, Pitman H, Oien K, *et al.* Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *Journal of Pathology J Pathol* 2019;**249**:143–50. doi:10.1002/path.5310

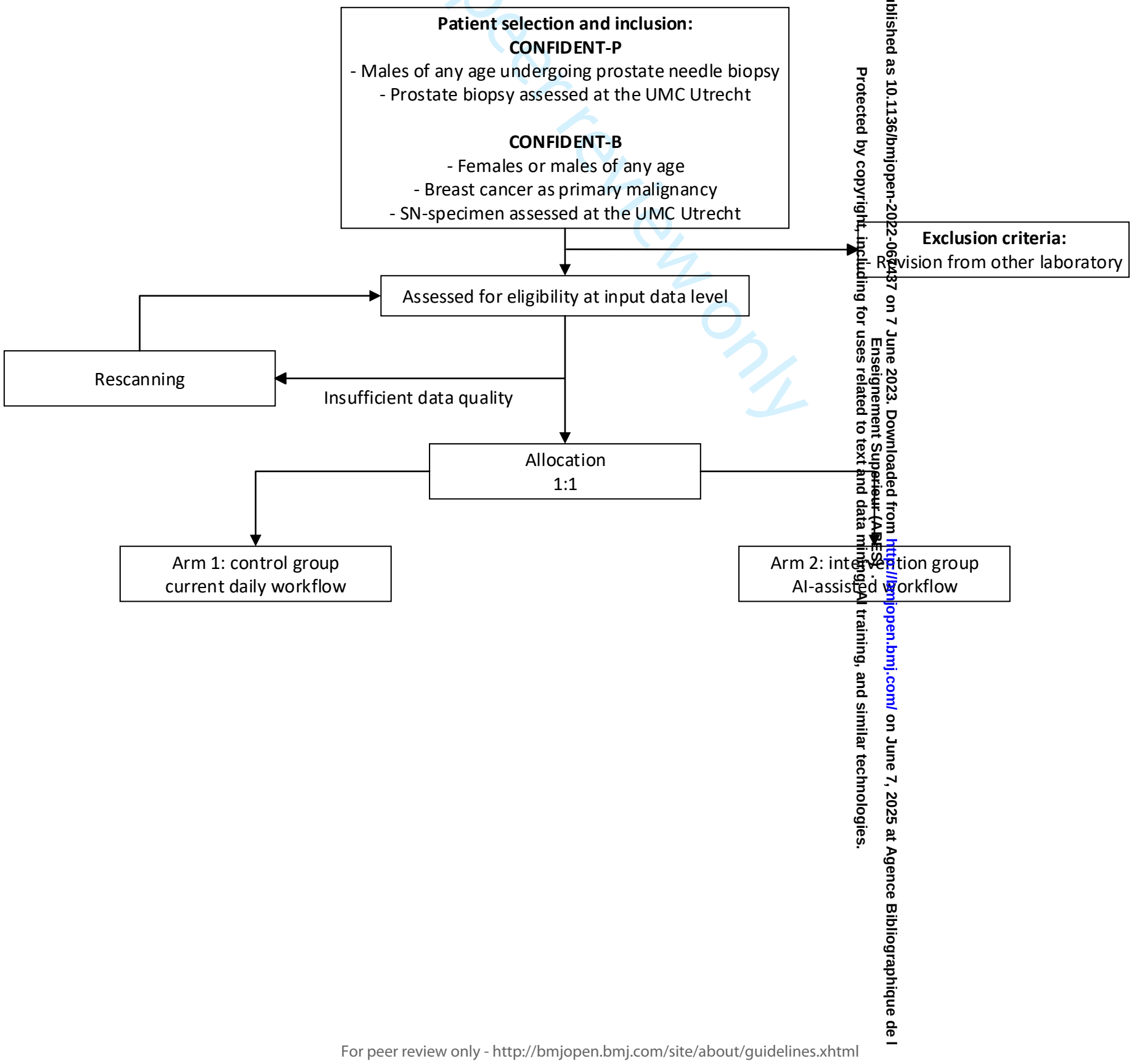
31 Kuijpers CCHJ, Sluijter CE, von der Thüsen JH, *et al.* Interlaboratory variability in the histologic grading of colorectal adenocarcinomas in a nationwide cohort. *American Journal of Surgical Pathology* 2016;**40**:1100–8. doi:10.1111/his.12923

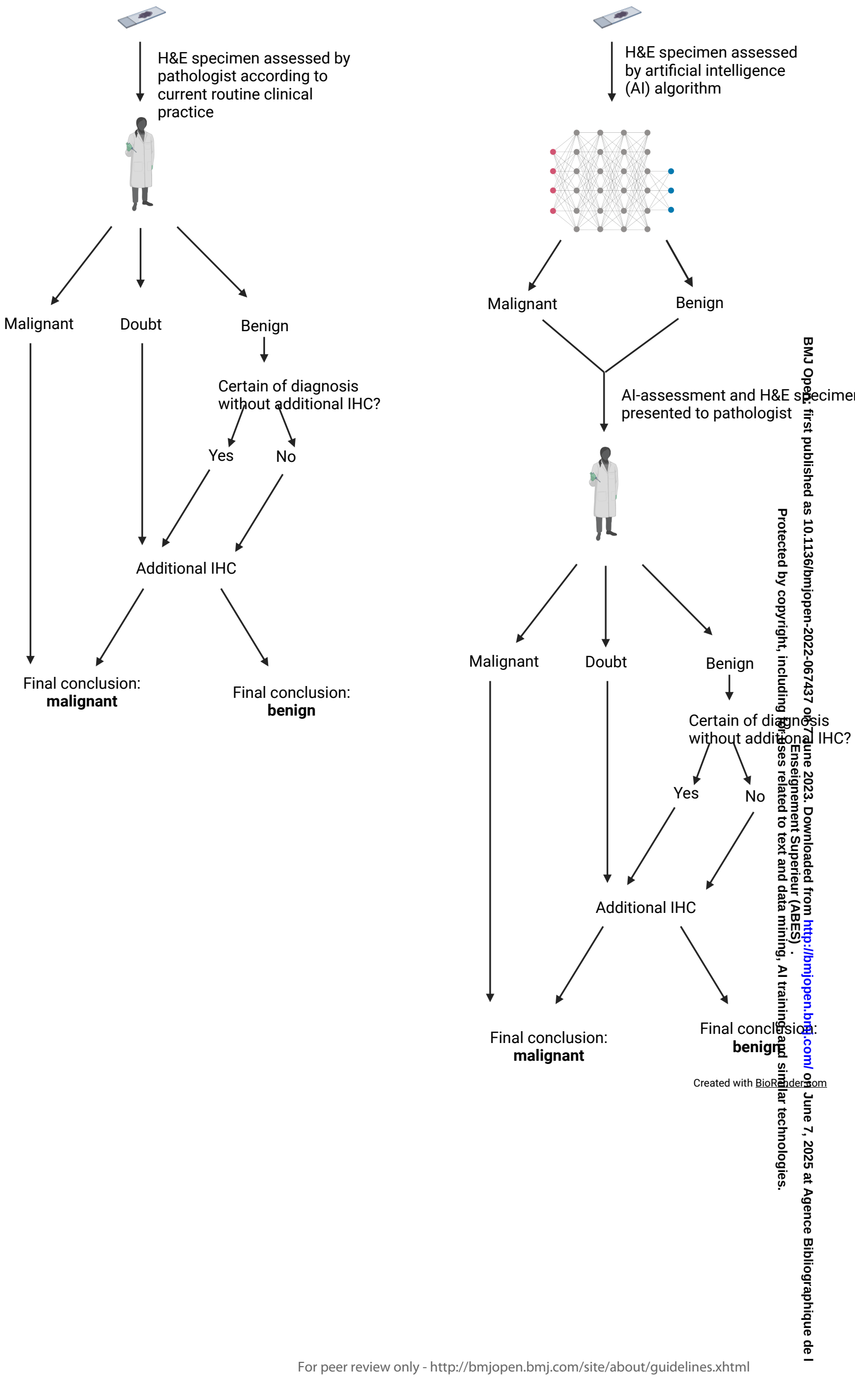
32 Kuijpers CCHJ, Sluijter CE, von der Thüsen JH, *et al.* Interlaboratory variability in the grading of dysplasia in a nationwide cohort of colorectal adenomas. *Histopathology* 2016;**69**:187–97. doi:10.1111/his.12923

33 Flach RN, Willemse P-PM, Suelmann BBM, *et al.* Significant Inter- and Intralaboratory Variation in Gleason Grading of Prostate Cancer: A Nationwide Study of 35,258 Patients in The Netherlands. *Cancers (Basel)* 2021;**13**. doi:10.3390/cancers13215378

34 van Dooijeweert C, van Diest PJ, Willems SM, *et al.* Significant inter- and intra-laboratory variation in grading of ductal carcinoma in situ of the breast: a nationwide study of 4901 patients in the Netherlands. *Breast Cancer Res Treat* 2019;**174**:479–88. doi:10.1007/s10549-018-05082-y

35 van Dooijeweert C, van Diest PJ, Willems SM, *et al.* Significant inter- and intra-laboratory variation in grading of invasive breast cancer: A nationwide study of 33,043 patients in the Netherlands. *Int J Cancer* 2020;**146**:769–80. doi:10.1002/ijc.32330





# Reporting checklist for protocol of a clinical trial.

Based on the SPIRIT guidelines.

## Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the SPIRITreporting guidelines, and cite them as:

Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin J, Dickersin K, Hróbjartsson A, Schulz KF, Parulekar WR, Krleža-Jerić K, Laupacis A, Moher D. SPIRIT 2013 Explanation and Elaboration: Guidance for protocols of clinical trials. BMJ. 2013;346:e7586

| | | Reporting Item | Page Number |
|---|---|---|---|
| **Administrative information** | | | |
| Title | #1 | Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym | 1 |
| Trial registration | #2a | Trial identifier and registry name. If not yet registered, name of intended registry | na |
| Trial registration: data set | #2b | All items from the World Health Organization Trial Registration Data Set | na |
| Protocol version | #3 | Date and version identifier | 1 |
| Funding | #4 | Sources and types of financial, material, and other support | 1 |
| Roles and responsibilities: contributorship | #5a | Names, affiliations, and roles of protocol contributors | 1 |

| | | | |
|---|---|---|---|
| Roles and responsibilities: sponsor contact information | #5b | Name and contact information for the trial sponsor | na |
| Roles and responsibilities: sponsor and funder | #5c | Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities | na |
| Roles and responsibilities: committees | #5d | Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee) | na |
| **Introduction** | | | |
| Background and rationale | #6a | Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention | 3 |
| Background and rationale: choice of comparators | #6b | Explanation for choice of comparators | 3 |
| Objectives | #7 | Specific objectives or hypotheses | 4 |
| Trial design | #8 | Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, non-inferiority, exploratory) | 4 |
| **Methods: Participants, interventions, and outcomes** | | | |
| Study setting | #9 | Description of study settings (eg, community clinic, academic hospital) and list of countries where data will | 4-5 |

| | | | |
|---|---|---|---|
| | | be collected. Reference to where list of study sites can be obtained | |
| Eligibility criteria | #10 | Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists) | 5 |
| Interventions: description | #11a | Interventions for each group with sufficient detail to allow replication, including how and when they will be administered | 6 |
| Interventions: modifications | #11b | Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving / worsening disease) | na |
| Interventions: adherance | #11c | Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return; laboratory tests) | na |
| Interventions: concomitant care | #11d | Relevant concomitant care and interventions that are permitted or prohibited during the trial | na |
| Outcomes | #12 | Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended | 6, 8 |
| Participant timeline | #13 | Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure) | 6 |
| Sample size | #14 | Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations | 7 |
| Recruitment | #15 | Strategies for achieving adequate participant enrolment to reach target sample size | 5 |

| | | | |
|---|---|---|---|
| **Methods: Assignment of interventions (for controlled trials)** | | | |
| Allocation: sequence generation | #16a | Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions | 5 |
| Allocation concealment mechanism | #16b | Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned | na |
| Allocation: implementation | #16c | Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions | na |
| Blinding (masking) | #17a | Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how | na |
| Blinding (masking): emergency unblinding | #17b | If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial | na |
| **Methods: Data collection, management, and analysis** | | | |
| Data collection plan | #18a | Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. | 8 |

| | | | |
|---|---|---|---|
| | | Reference to where data collection forms can be found, if not in the protocol | |
| Data collection plan: retention | #18b | Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols | na |
| Data management | #19 | Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol | 8-9 |
| Statistics: outcomes | #20a | Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol | 7-8 |
| Statistics: additional analyses | #20b | Methods for any additional analyses (eg, subgroup and adjusted analyses) | 7-8 |
| Statistics: analysis population and missing data | #20c | Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation) | 7-8 |
| **Methods: Monitoring** | | | |
| Data monitoring: formal committee | #21a | Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed | 8 |
| Data monitoring: interim analysis | #21b | Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial | na |
| Harms | #22 | Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events | 8 |

| | | | |
|---|---|---|---|
| | | and other unintended effects of trial interventions or trial conduct | |
| Auditing | #23 | Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor | na |
| **Ethics and dissemination** | | | |
| Research ethics approval | #24 | Plans for seeking research ethics committee / institutional review board (REC / IRB) approval | 9 |
| Protocol amendments | #25 | Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC / IRBs, trial participants, trial registries, journals, regulators) | na |
| Consent or assent | #26a | Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32) | na |
| Consent or assent: ancillary studies | #26b | Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable | 8 |
| Confidentiality | #27 | How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial | 8 |
| Declaration of interests | #28 | Financial and other competing interests for principal investigators for the overall trial and each study site | 9 |
| Data access | #29 | Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators | 8-9 |
| Ancillary and post trial care | #30 | Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation | na |
| Dissemination policy: trial results | #31a | Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the | 9 |

| | | | |
|---|---|---|---|
| | | public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions | |
| Dissemination policy: authorship | #31b | Authorship eligibility guidelines and any intended use of professional writers | 9 |
| Dissemination policy: reproducible research | #31c | Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code | 9 |
| **Appendices** | | | |
| Informed consent materials | #32 | Model consent form and other related documentation given to participants and authorised surrogates | 8 |
| Biological specimens | #33 | Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable | na |