QUADAS-2 tool reproduced with permission from the University of Bristol.[35]

First author surname and year of publication:

Name of first reviewer: *Name of second reviewer:*

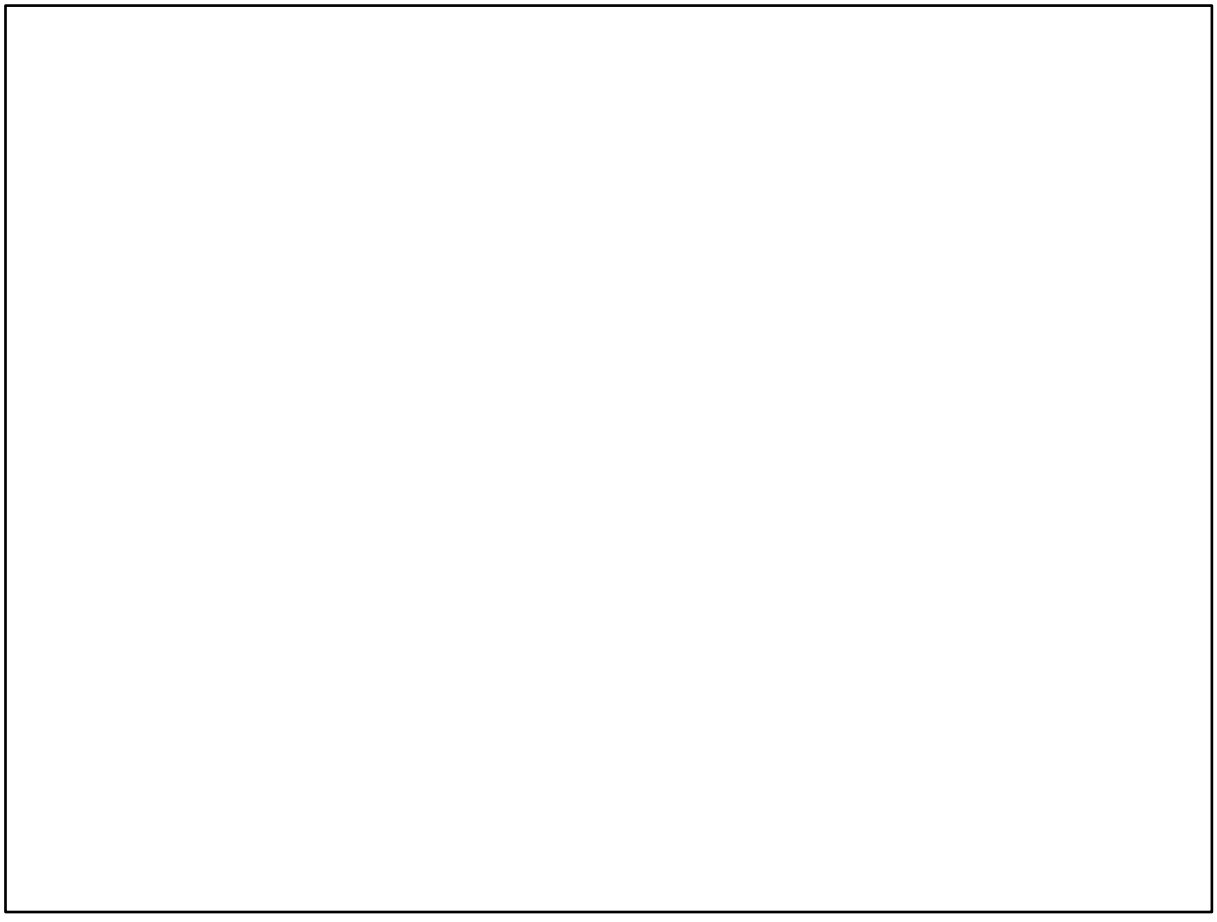
Date completed: *Date completed:*

Phase 1: State the review question:

What are the test accuracy, test failure rates, and time to diagnosis of IHC and MSI-based strategies for detecting Lynch syndrome in people who have a diagnosis of endometrial cancer?

Patients (setting, intended use of index test, presentation, prior testing):
Index test(s):
Reference standard and target condition:

Phase 2: Draw a flow diagram for the primary study



Phase 3: Risk of bias and applicability judgments

QUADAS-2 is structured so that 4 key domains are each rated in terms of the risk of bias and the concern regarding applicability to the research question (as defined above). Each key domain has a set of signalling questions to help reach the judgments regarding bias and applicability.

DOMAIN 1: PATIENT SELECTION	
A. Risk of Bias	
Describe methods of patient selection:	
+ Was a consecutive or random sample of patients enrolled?	Yes/No/Unclear
+ Was a case-control design avoided?	Yes/No/Unclear
+ Did the study avoid inappropriate exclusions?	Yes/No/Unclear
Could the selection of patients have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting)**:**	
Is there concern that the included patients do not match the review question?	**CONCERN: LOW/HIGH/UNCLEAR**

DOMAIN 2: INDEX TEST(S)

If more than one index test was used, please complete for each test.

A. Risk of Bias

Describe the index test and how it was conducted and interpreted:

+ Were the index test results interpreted without knowledge of the results of the reference standard?	Yes/No/Unclear
+ Were thresholds pre-specified?	Yes/No/Unclear
+ Were quality assurances measures in place?	Yes/No/Unclear
Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**CONCERN: LOW/HIGH/UNCLEAR**

DOMAIN 3: REFERENCE STANDARD

If more than one reference standard was used, please complete for each test.

A. Risk of Bias

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

+ Is the reference standard likely to correctly classify the target condition?	Yes/No/Unclear
+ Were the reference standard results interpreted without knowledge of the results of the index test?	Yes/No/Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**CONCERN: LOW/HIGH/UNCLEAR**

DOMAIN 4: FLOW AND TIMING **A. Risk of Bias**	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	
Describe the time interval and any intervention between index tests(s) and reference standard:	
+ Did all patients receive a reference standard?	Yes/No/Unclear
+ Did all patients receive the same reference standard?	Yes/No/Unclear
+ Were all patients included in the analysis?	Yes/No/Unclear
Could the patient flow have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**

DOMAIN 5: ROLE OF SPONSOR **A. Risk of Bias**	
+ Did the funding source/sponsor play no role in design of study, interpretation of results and publication?	Yes/No/Unclear
Could the funding source have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**

Modified QUADAS-2 and guidance notes

Parts of this section have been reproduced with permission from the University of Bristol.[36]

For each of the domains, risk of bias should be rated as 'low' if all signaling questions are answered with 'yes'. If one or more signaling question is answered with 'no' the risk of bias should be rated as 'high'. If none of the signaling question is answered 'no' and at least one question is answered with 'unclear', the risk of bias should be judged 'unclear'.

Domain 1: Patient selection

A. Risk of bias

Guidance:

Was a consecutive or random sample of people with endometrial cancer enrolled?

This question should only be answered 'yes' if the study clearly states that people with endometrial cancer were recruited consecutively or randomly. This question should be answered 'no' if the study clearly states that people with endometrial cancer were not recruited consecutively or randomly.

Was a case-control design avoided?

We would expect prospective cohort designs. Therefore, if the study is a case-control study this question should be answered with 'no'.

Did the study avoid inappropriate exclusions?

If the study excludes potential participants inappropriately (e.g. because they are difficult to diagnose, have had a previous or have a synchronous malignancy, or because of their age) or if >10% of participants are excluded either with or without specifying reasons, the exclusions should be considered as inappropriate. This cut-off has been determined pragmatically.

B. Concerns regarding applicability

Guidance:

For applicability concerns to be low, the study participants should be comparable to the eligible UK population (e.g. in terms of age range and ethnicity). If testing for Lynch syndrome in people with endometrial cancer is introduced in the UK, no age restrictions are

anticipated. Therefore, any study that limits participants by age will be considered to have high applicability concerns.

The setting of the study might have an impact on the applicability of the study results to general practice in terms of feasibility, if the equipment or standards of the study setting are unlikely to be met by the routine laboratory carrying out the tests in clinical practice in the UK. Some of the technologies used in the studies might not be feasible to be carried out in routine laboratories. It needs to be decided how applicable the results of these studies are to routine practice but also whether the index test is likely to be carried out in routine laboratories or in a few specialised centers.

Domain 2: Index test

The main sources of bias introduced by conducting and interpreting the index test are blinding, defining the threshold, the subjectivity of tests, and lack of quality assurance. If the reference standard is carried out before the index test (e.g. in case-control studies) it is important to blind personnel to the results of the reference standard. The QUADAS-2 tool requires a threshold to be pre-specified in the methods in order to avoid adjustment of the threshold according to the test outcome. There is some subjectivity involved in interpreting immunohistochemistry results. Tumours that show an absence of nuclear staining are rated as being 'negative' for the expression of the particular protein(s). Tumours that show nuclear staining are rated as being 'positive' for the expression of the particular protein(s). However, the amount and intensity of staining is important, and different studies have used different amounts and intensity of staining to indicate positive/negative expression of proteins. Factors that can affect the conduct of testing and accuracy of interpretation include pathologist experience, adequacy of biopsy sample (tumour content of >30% has been suggested for MSI and MHL1 promoter hypermethylation testing, e.g. to avoid false negative results), and the type of control sample (e.g. blood or normal tissue from matched-control).

A. Risk of bias

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

The studies need to report blinding clearly in order to answer this question with 'yes'.

Were thresholds pre-specified?

For this question to be answered 'yes' the study needs to mention the threshold used (e.g. microsatellite instability-based testing rated as 'positive' if 30% or more microsatellite markers show instability; immunohistochemistry rated as negative if unequivocal absence of staining or if <10% of the tumor is stained) and clearly state that it was specified before the start of the study. If the study reports adjustment to the threshold and reports results according to adjusted thresholds this question should be answered with 'no'.

Were quality assurances measures in place?

For this question to be answered 'yes' studies should indicate that the laboratories performing the index tests participate in an accredited quality assessment/control scheme, e.g. UK-National External Quality Assessment Scheme, Nordic immunohistochemical Quality Control, Clinical Laboratory Improvement Amendments programme. This question should be answered 'no' for studies that do not mention quality assurance being in place.

B. Concerns about applicability

Concerns about applicability will be low for studies that conduct and interpret index tests in accordance to best practice guidelines and via laboratories that are participating in quality assurance programmes. Applicability concerns will be high for studies not adhering to these standards, for example those that use experimental/research-only methods for index testing.

Domain 3: Reference standard

There is no single test that is used to identify all cases of Lynch syndrome. Lynch syndrome is diagnosed on the basis of constitutional mutations (i.e. mutations that are present in every cell) in MMR genes. This involves sequencing to detect point mutation, small insertions or deletions in these genes, and techniques such as multiplex ligation-dependent probe amplification to detect larger structural changes (i.e. deletions, duplications or rearrangements) to genetic sequences that could be missed by sequencing alone.

A. Risk of bias

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

This question will be answered with ‘yes’ for studies that use (1) sequencing to detect point mutations in combination with (2) multiplex ligation-dependent probe amplification, next-generation copy number, long-range PCR or targeted array comparative genome hybridisation to detect larger rearrangements or for dosage analysis. The process of conducting testing for constitutional mutations and interpretation of mutations should be carried out in accordance to best practice guidelines (e.g. Association for Clinical Genetic Services Best Practice Guidelines for Genetic Testing and Diagnosis of Lynch Syndrome, American College of Medical Genetics and Genomics Standards and Guidelines for Clinical Genetics Laboratories) in appropriately accredited laboratories (e.g. according to the UK Accreditation Service, the Clinical Laboratory Improvement Amendments). If studies use other reference standards or do not use methods to detect both point mutations and detect larger structural abnormalities together the question should be answered as ‘no’. If studies do not report the testing standard performed and the accreditation of the testing laboratories, n the question should be answered as ‘unclear.

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

This question should be answered with ‘yes’ if blinding of the index result is explicitly stated.

B. Concerns about applicability

Applicability concerns for the reference standard will be low if Lynch syndrome is diagnosed by germline testing for constitutional mutations in MMR genes by sequencing (as a minimum). It will be high if any other non-applicable reference standard (see protocol) is used (in the absence of sequencing), or if >10% of those reported as having Lynch syndrome have genetic variants of unknown clinical significance, Lynch-like syndrome, or ‘presumed’ Lynch syndrome (other terms are used and need to be assessed on a case-by-case basis) and their data cannot be excluded from our analyses. This threshold has been determined pragmatically.

Domain 4: Flow and Timing

A. Risk of bias

Did all participants receive a reference standard?

This question can only be answer with 'yes' if the all participants undergo germline testing using at least one of the reference standards mentioned above. The question should be answered with 'unclear' if the study provides no information on how controls were identified in case-control studies and risk of bias should be classed as 'high'.

Did all participants receive the same reference standard?

This question should be answered with 'no' if people received different reference standards, including if people with a positive tumour test result received a different reference standard to people with a negative tumour test result. This question should be answered with 'unclear' if a list of reference standards is given but no report is made of which people received which reference standard(s).

Were all participants included in the analysis?

If inconclusive or intermediate results or participants lost to follow up are not considered in the analysis the question should be answered with 'no' and the risk of bias considered 'high'. If studies report a clinical experience and base test accuracy estimates on interim results and not all people were followed up, the question should be answered with 'no' and the risk of bias should be classed as 'high'.

Domain 5: Role of sponsor

Studies that are sponsored by companies that manufacture the index tests might be biased if the company has influence on the study design, conduct, interpretation of results and decision to publish.

A. Risk of bias

Did the funding source/sponsor play <u>no role</u> in the design of study, interpretation of results, and publication?

The study needs to clearly state that sponsors played no role in order to answer this question with ’yes’. Equally, to answer the question with ‘no’ the study needs to clearly state sponsor involvement.