KCE REPORT 409S
GOOD CLINICAL PRACTICE

Federaal Kenniscentrum voor de Gezondheidszorg
Centre Fédéral d'Expertise des Soins de Santé
Belgian Health Care Knowledge Centre

METHODS GUIDELINE FOR THE DIAGNOSIS, TREATMENT, AND FOLLOW-UP OF CUTANEOUS MELANOMA IN SECONDARY CARE

SUPPLEMENT



VICKY JESPERS, ELENA COSTA, CÉLINE RICOUR, LEEN VERLEYE

2025 www.kce.fgov.be

COLOPHON

Title:	Methods guideline for the diagnosis, treatment, and follow-up of cutaneous melanoma in secondary care – Supplement
Authors:	Vicky Jespers (KCE), Elena Costa (KCE), Céline Ricour (KCE), Leen Verleye (KCE)
Programme manager:	Els Van Bruystegem (KCE)
Information specialist:	Nicolas Fairon (KCE)
Survey development:	Carl Devos (KCE)
External experts, GDG-members:	Jean-François Baurain (Cliniques universitaires Saint-Luc, UCLouvain), Francesca Maria Bosisio (UZ Leuven, KU Leuven), Lieve Brochez (UZ Gent), Astrid De Meulenaere (AZ Maria Middelares), Sam Dekeyser (AZ Maria Middelares), Ines Deleu (Belgian Society of Medical Oncology (BSMO)), Anne-Leen Deleu (Hôpital Universitaire de Bruxelles (HUB), Institut Jules Bordet), Diane Denys (Melanoompunt), Annie Drowart (Institut Jules Bordet), Marjan Garmyn (UZ Leuven), Evelyne Harkemanne (Cliniques universitaires Saint-Luc), Marc Haspeslagh (Dermpat), Niels Horst (Universitair Ziekenhuis Antwerpen (UZA)), Celine Jacobs (UZ Gent), Jolien Kessels (Vrije Universiteit Brussel (VUB), UZ Brussel), Vibeke Kruse (VITAZ), Mireille Langouo Fontsa (Institut Jules Bordet, (HUB)), Julie Lemahieu (Dermpat), Bart Neyns (UZ Brussel), Annemie Rutten (GZA Sint-Augustinus), Vasiliki Siozopoulou (Cliniques universitaires Saint-Luc, UCLouvain)
Stakeholders:	Marianne Bosman (Melanoompunt), Els Meerbergen (Kom op tegen Kanker), Diane-Estelle Ngatchou-Djomo (Ligue des usagers des services de santé (LUSS)), Arjen Nikkels (CHU Salt Tilman, Université de Liège (ULiège)), Stefan Teughels (Domus Medica), Liesbet Van Eycken (Kankerregister)
External validators:	Trudy Bekkering (Belgisch Centrum voor Evidence Based Medicine (CEBAM), Academisch Centrum voor Huisartsengeneeskunde (ACHG), KU Leuven), Dirk Grünhagen (Erasmus MC Kankerinstituut, the Netherlands), Sofie Wilgenhof (Antoni van Leeuwenhoek ziekenhuis, the Netherlands)
Acknowledgements:	We would like to thank the survey participants for their contribution.
Reported interests KCE members:	KCE is a federal institution funded by INAMI/RIZIV, by the Federal Public Service of Health, Food chain Safety and Environment, and by the Federal Public Service of Social Security. KCE's mission is to advise policymakers on decisions relating to health care and health insurance on the basis of scientific and objective research. It is expected to identify and shed light on the best possible solutions, in the context of an accessible, high-quality health care system with due regard for growing demand and budgetary constraints. KCE has no interest in companies (commercial or non-commercial, i.e. hospitals and universities), associations (e.g. professional associations, unions), individuals or organisations (e.g. lobby groups) that could be positively or negatively affected (financially or in any other way) by the implementation of the recommendations. All experts involved in the writing of the report or the peer-review process completed a declaration of interest form. Information on potential conflicts of interest is published in the colophon of this report. All members of the KCE Expert Team make yearly declarations of interest and further details of these are available upon request.
Reported interests:	All experts and stakeholders consulted within this report were selected because of their involvement in cutaneous melanoma. Therefore, by definition, each of them might have a certain degree of conflict of interest to the main topic of this report. **Membership of a stakeholder group on which the results of this report could have an impact:** Francesca Maria Bosisio (Dermatopathologen - Dermatopathology working group of the Belgian Society of Pathology), Ines Deleu (BSMO), Diane Denys (Association melanoma patients, Melanoompunt

vzw), Julie Lemahieu (Pathologists), Diane-Estelle Ngatchou-Djomo (Associations de patients melanoma membres de la LUSS)

Owner of subscribed capital, options, shares or other financial instruments: Julie Lemahieu (Parts of Dermpat, labo Pathology)

Fees or other compensation for writing a publication or participating in its development: Lieve Brochez (Co-author of the 2024 update of the melanoma guideline by EADO-ESMO-EORTC. Meeting Rome November 2023 stay paid for by EADO; no other compensations), Ines Deleu (Investigator meetings Pierre Fabre, Sun), Kruse Vibeke (Article breast cancer for BJMO + video TNBC. Podcast melanoma for Pierre Fabre Oncology. Short scientific report SCC for Sanofi. Participation in article about melanoma for MSD)

Participation in scientific or experimental research as an initiator, principal investigator or researcher: Francesca Maria Bosisio (Melanoma researcher spatial multiomics technologies), Lieve Brochez (Group leader Skin Cancer Research Institute Ghent: scientific research melanoma (investigator-initiated, not industry sponsored), Ines Deleu (Regeneron-study with Fianlimab adjuvant and metastatic), Annie Drowart (Study CA209-8RX BMS), Evelyne Harkemanne (Experimental study diagnosis and characterization melanoma by electron paramagnetic resonance), Celine Jacobs (PI various studies regarding melanoma patients (phase 2 and phase 3 studies). These commercial studies have no influence on the development of guidelines as these products are still being researched into their effectiveness compared to the standard of care. They are not yet available as current practice. Also, the content discussed via the KCE is not discussed with the study team or anyone from the pharmaceutical industry. Current studies: V940-001 study: "A Phase 3, Randomized, Double-Blind, Placebo and Active-Comparator-Controlled Clinical study of Adjuvant V940 (mRNA-4157) Plus Pembrolizumab Versus Adjuvant Placebo Plus Pembrolizumab in Participants With High-Risk Stage II-IV Melanoma". Ongoing study, but no longer active inclusion: *IO102-IO103. Past studies* “Monette study*CA209-8RX study* CA224-047 study), Julie Lemahieu (Modest and organic investigator in dermatopathology of melanocytic lesions), Bart Neyns (PI for studies with sponsor Novartis, MSD, BMS, Pierre Fabre), Annemie Rutten (PI in clinical trials melanoma with MSD and BMS as sponsor), Kruse Vibeke (PI in two melanoma studies, SI in different studies), Sofie Wilgenhof (Principal investigator: CA224020 study (ClinicalTrials.gov number, NCT01968109), R3767-ONC-2011 study (NCT05352672), E2139 (NCT05270044), vorinostat study (NCT02836548))

A grant, fees or funds for a member of staff or another form of compensation for the execution of research described above: Francesca Maria Bosisio (FWO Senior Clinical Researcher bursary), Lieve Brochez (Funding Kom op tegen Kanker immuno-oncology project: final phase; Baekeland grant HBC.2019.2572: New classification and image recognition of melanoma risk of nevi based on dermoscopy, targeted microscopy and clinico-epidemiological data (2019-2024)), Evelyne Harkemanne (FNRS personal research post-doctoral specialist part-time), Jolien Kessels (Sub-investigator: NCT05933577), Julie Lemahieu (Baekeland mandaat Veronique Clauwaert - classification of flat melanocytic lesions), Sofie Wilgenhof (CAPABLE: Pilot study of the eHealth application Cancer Patients Better Life Experience: EU Horizon 2020 (No 875052), CARE-1: Optimizing Treatments for Metastatic Renal Carcinoma: EU Horizon: EU (101104801))

Consultancy or employment for a company, an association or an organisation that may gain or lose financially due to the results of this report: Jolien Kessels (Consultancy: Bioderma, L’Oreal), Annemie Rutten (Advisory board MSD, BMS, Pierre Fabre), Kruse Vibeke (consultancy fees from Novartis, MSD, BMS and Pierre Fabre oncology for the participation in meetings. My institution also received sponsoring from different pharmaceutical companies for the organisation of scientific meetings), Sofie Wilgenhof (Consultancy IPSEN (paid to AVL/NKI Institute) Spoke at MSD educational event (paid to AVL/NKI Institute))

Payments to speak, training remuneration, subsidised travel or payment for participation at a conference: Ines Deleu (EADO, ESMO), Annie Drowart (Inscription EADO 2024 BMS), Niels Horst (EADO meeting with Sun Pharma payment; I recently spoke at YADE, but this was about skin infections and colored skin), Celine Jacobs (05/2024: BMS organized symposium: speaker fee* 12/2023: MSD Advisory Board,* 11/2022: Novartis meeting,* 4/2024: EADO conference (MSD),* 6/2024: ASCO conference entrance ticket online (Novartis),* 10/2023: online ticket ESMO conference (MSD),* 04/2022: EADO (Sanofi)), Jolien Kessels (Speaker for fee: Bristol Myers Squibb, EMGEN, Janssen, training support/travel allowance: MSD, Astrazeneca), Mireille Langouo Fontsa (Pierre Fabre: payment of conference fees EADO, BMS: Meeting speaker), Bart Neyns (MSD, BMS, Pierre Fabre, Novartis), Annemie Rutten (Travel support MDS, payment to speak BMS)

Presidency or accountable function within an institution, association, department, or other entity on which the results of this report could have an impact: Francesca Maria Bosisio (Vice-President of the Dermatopathology working group of the Belgian Society of Pathology), Lieve Brochez (Board member Belgian Association of Dermato-Oncology (BADO)), Niels Horst (Board member of Belgian Mohs Group. Doesn't seem like a conflict of interest to me because we don't use Mohs surgery for pigment lesions), Julie Lemahieu (Managing director Dermpat), Sofie Wilgenhof (Member of the guideline committee (Dutch Internists Association) Melanoma and Skin Malignancies)

Layout:	Ine Verhulst

Disclaimer:

- **The external experts were consulted about a (preliminary) version of the scientific report. Their comments were discussed during meetings. They did not co-author the scientific report and did not necessarily agree with its content.**
- **Subsequently, a (final) version was submitted to the validators. The validation of the report results from a consensus or a voting process between the validators. The validators did not co-author the scientific report and did not necessarily all three agree with its content.**
- **Finally, this report has been approved Choose an item. by the Executive Board.**
- **Only the KCE is responsible for errors or omissions that could persist. The policy recommendations are also under the full responsibility of the KCE.**

Publication date:	06 November 2025
Project number:	2018-05-GCP
Domain:	Good Clinical Practice (GCP)
MeSH:	Melanoma, Cutaneous Malignant; guideline, Belgium, Health Personnel
NLM Classification:	WR500
Language:	English
Format:	Adobe® PDF™ (A4)
Legal depot:	D/2025/10.273/47
ISSN:	2466-6459

Copyright: KCE reports are published under a “by/nc/nd” Creative Commons Licence http://kce.fgov.be/content/about-copyrights-for-kce-publications.

How to refer to this document?

Jespers V, Costa E, Ricour C, Verleye L. Methods guideline for the diagnosis, treatment, and follow-up of cutaneous melanoma in secondary care. Good Clinical Practice (GCP) Brussels: Belgian Health Care Knowledge Centre (KCE). 2025. KCE Reports 409S. https://doi.org/10.57598/R409CS.

This document is available on the website of the Belgian Health Care Knowledge Centre.

TABLE OF CONTENTS

1 GENERAL ASPECTS OF A KCE GUIDELINE 4
1.1 FUNDING AND DECLARATION OF INTEREST 4
1.2 STATEMENT OF INTENT 4
1.3 TARGET USERS OF A GUIDELINE 4
1.4 STAKEHOLDERS, EXPERTS, PATIENT INVOLVEMENT 4
1.4.1 The guideline development group 4
1.4.2 Stakeholders 5
1.4.3 Patient involvement 5
2 METHODOLOGY 5
2.1 SCOPING PHASE 5
2.2 DEVELOPMENT OF RESEARCH QUESTIONS 6
2.3 THE GRADE-ADOLOPMENT APPROACH 6
2.4 EVIDENCE REVIEW 8
2.4.1 Selection and quality assessment of systematic reviews 9
2.4.2 Selection and quality assessment of primary studies 10
2.4.1 Summarising the evidence 10
2.4.2 The quality of evidence 10
2.5 FROM EVIDENCE TO RECOMMENDATIONS 12
2.6 GOOD PRACTICE STATEMENTS AND CLINICAL PRACTICE GUIDANCE 13
2.7 EXTERNAL REVIEW AND VALIDATION 14
2.7.1 Stakeholder review 14
2.7.2 Validation by external experts 14
2.7.3 Validation by the Belgian Centre for Evidence-Based Medicine 15
2.8 DISSEMINATION AND IMPLEMENTATION 15
3 VALIDATED ASSESSMENT TOOLS 16
3.1 AGREE II – QUALITY GUIDELINES 16
3.2 AMSTAR 2 – QUALITY SYSTEMATIC REVIEWS 17
3.3 QUADAS-2 – RISK OF BIAS DIAGNOSTIC STUDIES 22
3.4 ROB-1 – RISK OF BIAS FOR RCTS 25
3.5 ROBINS-I – RISK OF BIAS IN NON-RANDOMISED STUDIES OF INTERVENTIONS 26

LIST OF FIGURES

Figure 1 – The Grading of Recommendations, Assessment, Development and Evaluations (GRADE)-ADOLOPMENT approach 8

LIST OF TABLES

Table 1 – A summary of the GRADE approach to grading the quality of evidence for each outcome 11
Table 2 – Downgrading the quality rating of evidence using GRADE 11
Table 3 – Strength of recommendation according to the GRADE system 12
Table 4 – Factors that influence the strength of a recommendation 12
Table 5 – Interpretation of strong and conditional recommendations 13
Table 6 – GRADE: developing good practice statements 14
Table 7 – AGREE II instrument 16
Table 8 – Evaluation criteria AMSTAR-2 for a systematic review 17
Table 9 – 4 domains of the QUADAS-2 assessment 22
Table 10 – Quality of diagnostic studies with QUADAS-2 assessment 23
Table 11 – Cochrane Collaboration's risk of bias tool 1 for assessing risk of bias 25

LIST OF ABBREVIATIONS

ABBREVIATION	DEFINITION
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
GDG	Guideline development group
KCE	Belgian Health Care Knowledge Centre
NIHDI	National Institute for Health and Disability Insurance (RIZIV – INAMI)
PICO	Participants–Interventions–Comparator–Outcomes
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies 2
RCT	randomised controlled trial
SR	systematic review

METHODOLOGY SUPPLEMENT

1 GENERAL ASPECTS OF A KCE GUIDELINE

1.1 Funding and declaration of interest

The Belgian Health Care Knowledge Centre (KCE) is a federal institution funded for the largest part by the National Institute for Health and Disability Insurance (NIHDI - RIZIV – INAMI), but also by the Federal Public Service Health, Food Chain Safety and Environment (FOD VVVL – SPF SPSCAE) and the Federal Public Service of Social Security (FOD SZ – SPF SS). The funders are not involved in the development of clinical recommendations. The development of clinical practice guidelines is part of the legal mission of the KCE. KCE has no interest in companies (commercial or non-commercial like for instance hospitals and universities), associations (e.g. professional associations, unions), individuals or organisations (e.g. lobby groups) that could be positively or negatively affected (financially or in any other way) by the implementation of a KCE guideline. All clinicians belonging to the Guideline Development Group, as well as the stakeholders and the experts involved in the peer-review process completed a declaration of interest form. Information on potential conflicts of interest is published in the colophon of the report related to the guideline. All researchers of the KCE also make yearly declarations of interest; further details are available upon request.

1.2 Statement of intent

Clinical guidelines are designed to improve the quality of health care and to decrease the use of unnecessary or harmful interventions. KCE guidelines are developed for use within the Belgian healthcare context.

The recommendations are not intended to indicate an exclusive course of action or to serve as a standard of care. Standards of care are determined on the basis of all the available clinical data for an individual case and are subject to change as scientific knowledge and technology advance and patterns of care evolve. Variations, which take into account individual circumstances, clinical judgement and patient choice, may also be appropriate.

The information in a KCE guideline is not a substitute for proper diagnosis and treatment, or for the provision of advice by an appropriate health professional. It is advised, however, that any significant deviation from a national guideline is fully documented in the patient's file at the time the relevant decision is taken.

1.3 Target users of a guideline

At KCE, a guideline is primarily developed for health professionals in secondary care.

In addition, the authors aim that the guideline will be used in academic teaching, as well as in continuous education programmes for medical professionals. A guideline may also be of interest to primary care physicians, patients and their families, and to policy makers.

1.4 Stakeholders, experts, patient involvement

1.4.1 The guideline development group

A KCE guideline is developed by KCE experts in collaboration with a multidisciplinary group of health professionals named a guideline development group (GDG). At least one patient representative is desirable but not mandatory. Other representatives that may want to be included by the KCE researchers depending on the research topic are representatives from NIHDI and health insurance(s). The group is finalised after an extensive search for experts in Belgium on the topic, for example, as authors on relevant publications, or leading or recognised experts in the field. Individuals or

universities are contacted and asked for their interest to participate. The roles assigned to the GDG are:

- Definition of the clinical questions with KCE and stakeholders
- For the review of the literature
  - Feedback on outcomes and thresholds for clinical relevance
  - Feedback on the selection of papers and identification of missed papers
  - Feedback on the quality of evidence, e.g. indirectness (applicability)
- Feedback on the content of the guideline and its recommendations
  - Feedback on health benefits, harms, certainty in the best available evidence, cost, feasibility and how this might affect the final recommendation
  - Adapt formulation to the local context where needed

1.4.2 Stakeholders

Stakeholders are representatives of the relevant professional societies, patients' representatives, or other end users of the guideline. Stakeholders are individuals or groups who are affected by a guideline's development and implementation. Stakeholders are not involved in the guideline development group. The stakeholders provide a patient perspective on the priority of the research questions during the scoping phase. They review the draft or final recommendations for clarity of the wording, feasibility and acceptability for patients, socio-cultural acceptability, feasibility for the health system, health equity, and give comments on the methods to facilitate uptake of recommendations. The stakeholders' advice is independent from the GDG. The stakeholders group is finalised after an extensive search in Belgium on the topic, for example, patient organisations or professional organisations. Individuals or representatives of organisations are contacted and asked for their interest to participate.

1.4.3 Patient involvement

At KCE, the aim is to identify a representative mandated to speak for a patients' association involved in the topic of the guideline. If a dedicated patients' association does not exist, the umbrella patients' association of the different federal states are contacted. Again, if no candidates are identified a patient(s) will be considered on individual suitability for each project.

2 METHODOLOGY

A guideline is developed based on scientific evidence in accordance with the KCE guideline development methodology which is summarised in a process Book (this Book is currently been redeveloped and a new version will be available last quarter 2025 to include new GCP methodology for example developments in the GRADE methodology). The guideline process includes a scoping phase to define the topics, the development of the research question(s), an evidence review for each question (literature review, quality appraisal, analysis and summary, grading of the retrieved evidence), and the formulation of recommendations by applying the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) methodology.

2.1 Scoping phase

For defining the final scope and priority research questions experts and stakeholders are consulted. This can be through interviews, an (online) survey with stakeholders / experts, or at a meeting. The highest ranked research questions are selected based on feasibility and available resources.

In preparation of the scoping discussions, a systematic review for recent guidelines is performed and documented:

- searches of e.g. OVID Medline, the TRIP database and the G-I-N library database,
- searches of websites of scientific knowledge and professional organisations e.g.
  - Belgium: Cebam Digital Library for Health (CDLH vzw), Domus Medica
  - Europe: the National Institute for Health and Care Excellence (NICE) England, SIGN - Health Improvement Scotland, Federatie Medisch Specialisten Nederland, die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) Leitlinienregister for Germany, Haute Autorité de Santé (HAS) for France.
- For guidelines in oncology, examples are
  - European Society for Medical Oncology
  - The United States: the National Cancer Institute at NIH, the American Society of Clinical Oncology (ASCO), the Society for Immunotherapy of Cancer, the U.S. Preventive Services Task Force.
  - Canada: Cancer Care Alberta, Ontario Health Care.
  - Australia: Cancer Council Australia, the Royal Australian College of General Practitioners

The identified guidelines are then appraised with the AGREE II instrument.(AGREE Next Steps Consortium 2017 ,Brouwers 2010) The focus is on scoring the 8 aspects of the domain assessing the 'rigour of development'. This domain of the AGREE II instrument is the most decisive part to sort the guidelines during the scoping phase. Scoring is performed by one (or two) researcher(s).

During the scoping phase it may become clear that one or more research questions have been addressed by a recent guideline. The KCE researchers may thus decide to start from the identified source guideline of relevance and adapt or adopt certain recommendations.(Klugar 2024)

2.2 Development of research questions

Once the scope, the aim and objective of the guideline are clear, the final clinical research questions are developed based on the identified topics.

- The research questions are developed in collaboration with the members of the GDG.
- The Participants–Interventions–Comparator–Outcomes (PICO) framework is defined for each question. A PICO table can include the following items (depending on the type of review): Objective; Review questions; Population; Intervention; Comparator; Reference standard; Target condition; Outcomes (critical, important, cost, other); Study design; and/or Subgroup analyses.

2.3 The GRADE-ADOLOPMENT approach

The GRADE-ADOLOPMENT approach suggests to adopt or adapt recommendations from existing recent evidence-based guidelines that fulfil certain criteria for credibility and rigour of development.

If a recent evidence-based guideline, developed following rigorous methods is available, adoption or adaptation of the recommendations can be considered to avoid duplication of work and save on time and resources. 'Adoption' means the use of an existing trustworthy recommendation without modifications and without providing information on implementation for a specific situation. 'Adaption' means the use of an existing trustworthy recommendation with modifications such as a change in a specific element of the PICO, for example, the age of the population, or the certainty of evidence (when new evidence is identified), and may provide additional information on the local conditions for implementation.(Schünemann 2017)

Both adoption and adaptation of recommendations starts with the identification and critical appraisal of published guidelines that could serve as source.(Klugar 2024)

For the critical appraisal of the published guidelines, we focus on the rigour of development (domain 3 of the AGREE II checklist), more specifically on the methods used to search evidence, the criteria to select evidence and the reporting on the results (benefits and harms) and the link between the recommendations and the supporting evidence. To be included, guidelines ideally need a good score (for example more than 70% but this it topic depended) on the AGREE II instrument's domain 'rigour of development.(Klugar 2024)

In a next step, it has to be assessed whether the systematic review on which a recommendation is based needs updating. This decision can be guided by a representative sample of stakeholders, *in casu* the guideline development group. An additional systematic review may be needed to collect evidence on selected outcomes that were not included in the source guideline.(Klugar 2024)

Once a summary of the supporting evidence regarding a research question is complete, recommendations are derived from the available evidence taking the following factors into account:

- Balance between clinical benefits and harms
- Quality of the evidence
- Patient preferences and values
- Cost and resource allocation

The considerations that are reported by the source guideline are taken into account when available and adapted to the Belgian context. If the GDG agrees with the reasoning of the source guideline, the recommendation is adopted. If adaptations are needed because of new evidence or because of the Belgian context, the considerations of the GDG and possible changes to the recommendations are reported. Also, newly developed evidence to decision frameworks are reported when applicable.(Klugar 2024)

The decision to adopt or adapt can be taken seperately for each recommendation. If for a prioritised research question no credible recommendation was identified, a *de novo* recommendations is developed.

Figure 1 – The Grading of Recommendations, Assessment, Development and Evaluations (GRADE)-ADOLOPMENT approach

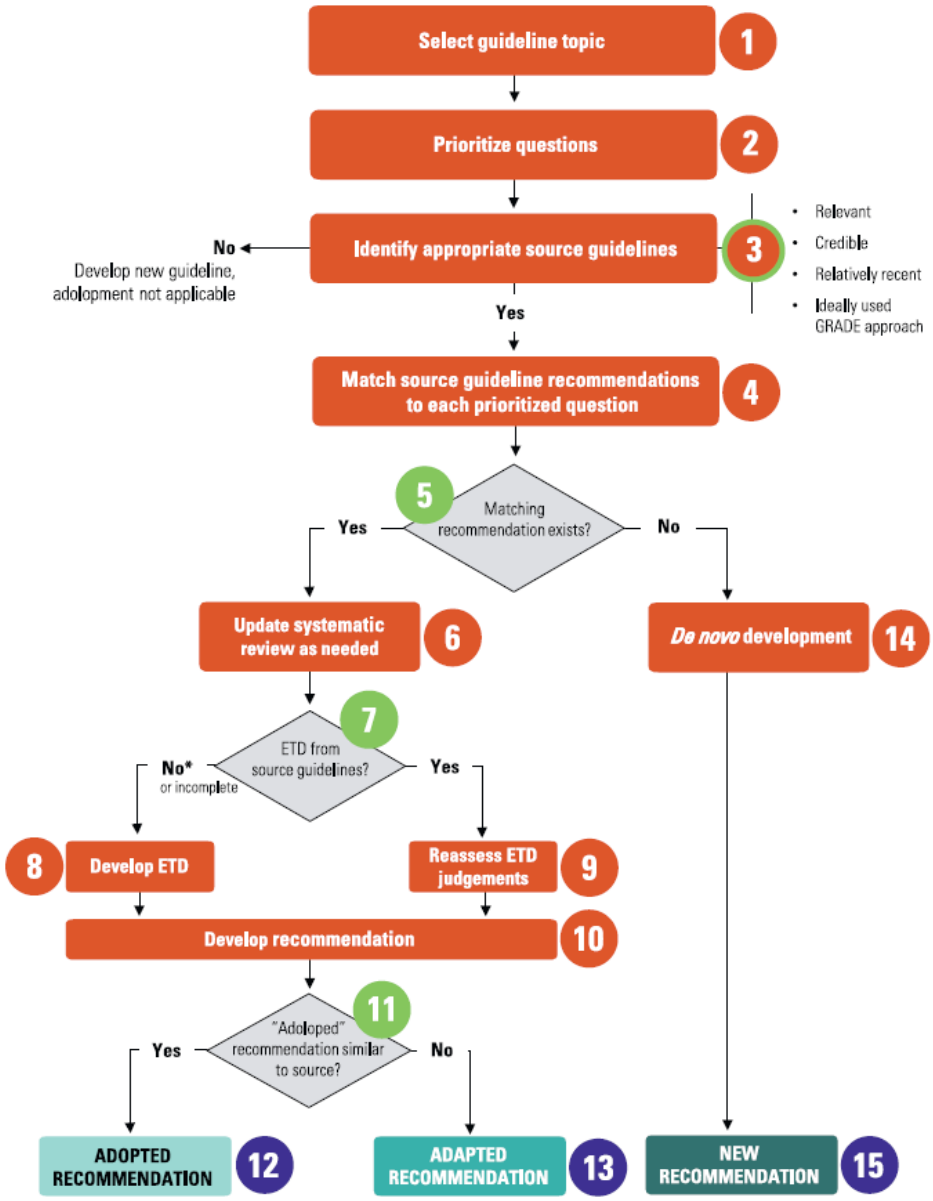


Figure copied from Klugar 2024(Klugar 2024)

2.4 Evidence review

For each research question a systematic literature review is set up by KCE experts (optionally the work can be outsourced) except in some cases when an adopt/adapt approach is used (see 2.3). An iterative approach can be followed to avoid duplication of work. For example, the evidence retrieved from a recent systematic review can be the starting point for an update. The search strategies and searches are documented in the report.

- the iterative approach starts with a search for recently published systematic review(s) (SR); then an update of the evidence synthesis is performed by searching for relevant primary studies published after the search date of the SR; If no SR is identified, a complete search for primary studies is performed.

- studies with a specific study design are identified e.g. SR, (diagnostic) randomised controlled trials (RCTs), observational (comparative) studies and/or prospective/retrospective diagnostic accuracy studies (in which the index test(s) and the reference standard test are applied to the same patients in a cross-sectional design). The design included is defined by the type of review (e.g. intervention vs diagnostic). The following study types are usually excluded: Narrative reviews; Case reports; Studies presented as conference abstract only or if no full text was available; and studies in which it is not possible to extract the required data.
- Databases to be used as a minimum for a search of a SR
  - Medline Ovid
  - At least one other electronic database
- Databases to be used as a minimum for a search of primary studies
  - Cochrane Library (Cochrane Reviews and Trials)
  - Medline Ovid
  - Embase
- The selection of publications and studies is performed by one (or two) researcher in two phases.
  - Phase one: screening of the titles and abstracts and excluding studies not fulfilling the inclusion criteria
  - Phase two: full text screening of selected studies. If no full-text is available, the study is not taken into account. Studies published in a language other than English, Dutch, French or German are usually not included. A table with reason for exclusion for each study is presented.

2.4.1 Selection and quality assessment of systematic reviews

A systematic review that would qualify as high quality has the following characteristics

- addresses at least one of the research questions;
- searches more than one database and is not restricted to MeSH terms;
- formulates clearly the in- and exclusion criteria;
- evaluates at least one of the selected (critical and important) outcomes;
- includes RCTs, comparative observational studies or observational cohort or cross-sectional studies regarding diagnostic test accuracy (depending on the research question);
- includes a risk of bias assessment of the main quality items with results reported for the individual studies;
- If more than one systematic review is identified for a particular research question, priority is on quality first and next recency (based on literature search date).

The Quality assessment is

performed by one researcher with the AMSTAR-2 (Assessing the methodological quality of systematic reviews) tool.(Shea 2017)

2.4.2 Selection and quality assessment of primary studies

A selected primary study that would qualify is

- a randomised intervention trial and/or (for example if non-randomised intervention trials are identified or the RCT evidence is very limited) a non-randomised, or comparative observational intervention study (see the guidance by Cochrane);
- a randomised diagnostic study or a study presenting diagnostic test accuracy measures (for example a cross-sectional study presenting diagnostic test accuracy) (see the guidance by Cochrane);
- addresses the correct patient populations and the comparisons in the PICO;
- evaluates at least one of the selected critical or important outcomes.

For the quality assessment of bias of

- diagnostic studies, the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) is used (Whiting 2011);
- RCTs, the Cochrane Collaboration's tool for assessing risk of bias is used (Higgins, Altman, on behalf of the Cochrane Statistical Methods Group and the Cochrane Bias Methods Group 2012);
- observational studies, the ROBINS-I (Risk of bias in non-randomised studies of interventions) tool is used.(Sterne 2016) This tool can be used when two or more interventions are compared in different sorts of non-randomised study designs, such as cohorts studies, case-control, before-after studies, interrupted-time series or quasi randomised studies. Similar to the Cochrane Risk of bias tool, the focus is on internal validity of a study (and not on generalisability or directness).

2.4.1 Summarising the evidence

The evidence on the patient-important outcomes for each research question is summarised in a GRADE evidence table and a summary of findings table. A GRADE evidence profile presents information about the 'body of evidence' (not the individual study), the judgements about the underlying quality of the evidence, the key statistical results, and the quality of evidence rating for each outcome (see GRADE handbook). The evidence report has conclusions and a level of evidence for each comparison and separate outcome.(Piggott 2022)

A summary of Findings table provides a summary of findings for each of the included outcomes and the quality of evidence rating for each outcome, without details of the judgements about the quality of evidence.

2.4.2 The quality of evidence

Applying GRADE facilitates a rigorous and transparent formulation of recommendations from evidence.(Guyatt, Oxman, Schünemann, Tugwell, Knottnerus 2011) The GDG should base its recommendation on the best available body of evidence related to the health care question. The GRADE approach to rating the quality of evidence begins with the study design (underlying quality of evidence e.g. trials or observational studies) and then addresses five reasons (risk of bias, imprecision, indirectness, inconsistency, and other domains) that increase the certainty.to possibly rate down the quality of evidence and three to possibly rate up the quality. The final overall quality of evidence in an effect estimate is defined in 4 categories; high, moderate, low or very low (Table 1)(Table 2).

Table 1 – A summary of the GRADE approach to grading the quality of evidence for each outcome

Source of body of evidence	Initial rating of quality of a body of evidence	Factors that may decrease the quality	Factors that may increase the quality	Final quality of a body of evidence
Randomised trials	High	1. Risk of bias 2. Inconsistency 3. Indirectness 4. Imprecision 5. other considerations (including publication bias and factors that increase the quality of evidence)	1. Large effect 2. Dose-response 3. All plausible residual confounding would reduce the demonstrated effect or would suggest a spurious effect if no effect was observed	High (⊕⊕⊕⊕) Moderate (⊕⊕⊕⊖) Low (⊕⊕⊖⊖) Very low (⊕⊖⊖⊖)
Observational studies	Low			

Source: GRADE handbook

Table 2 – Downgrading the quality rating of evidence using GRADE

Quality element	Reasons for downgrading
Risk of bias(Guyatt 2011)	Possible risk of bias: by lack of allocation concealment, lack of blinding, lack of intention-to-treat analysis, loss of follow-up and selective outcome reporting. Additionally, other limitations such as stopping early for benefit and use of unvalidated outcome measures can be taken into consideration.
Inconsistency(Guyatt 2011)	Downgrading the level of evidence for inconsistency of results: point estimates vary widely across studies, confidence intervals show minimal or no overlap, the statistical test for heterogeneity shows a low p-value or the I^2 is large. If large variability in magnitude of effect remains unexplained, the quality of evidence is rated down.
Indirectness / applicability(Guyatt 2011)	In case the trial population or the applied intervention differs significantly from the population or intervention of interest. Also, the use of surrogate outcomes or when studied interventions are not tested in a head-to-head comparison can lead to downgrading. For diagnostic studies, when there is concern that the included participants do not match the review question.
Imprecision(Guyatt 2011)	Examination of the 95% confidence interval (CI): Quality is rated down if clinical action would differ if the upper versus the lower boundary of the 95%CI represents the truth.
Publication bias(Guyatt 2011)	Publication bias is studied by analysis using funnel plots or searching of trial registries. Publication bias can also be suspected if results come from small, positive industry-sponsored trials only.

GRADE for guidelines considers the evidence across all outcomes and across studies for each research question. The evidence can thus be linked to each recommendation.

Guideline panels have to determine the overall quality of evidence across all the critical outcomes essential to a recommendation they make. Guideline panels usually provide a single grade of quality of evidence for every recommendation, but the strength of a recommendation usually depends on evidence regarding not just one, but a number of patient-important outcomes and on the quality of evidence for each of these outcomes.

When determining the overall quality of evidence across outcomes:

- Consider only those outcomes that are deemed critical;
- If the quality of evidence differs across critical outcomes and outcomes point in different directions — towards benefit and towards harm — the lowest quality of evidence for any of the critical outcomes determines the overall quality of evidence;
- All outcomes point in the same direction — towards either benefit or harm — the highest quality of evidence for a critical outcome that by itself would suffice to recommend an intervention determines the overall quality of evidence. However, if the balance of the benefits and downsides

is uncertain, then the grade of the critical outcome with the lowest quality grading should be assigned.

For an oncology guideline, in the evidence profiles, relative and absolute treatment effects are reported. Absolute treatment effects of an intervention can be calculated using the baseline risk as reported in a relevant clinical trial or as reported by the Belgian Cancer Registry. For example, to calculate absolute effects for time-to-event outcomes, the formulas published by Skoetz et al (2020) were used.(Skoetz 2020)

2.5 From evidence to recommendations

After the evidence is collected and summarised, and the GRADE criteria for rating the certainty of evidence applied, the recommendations are formulated.

Formulating a recommendation (even if the level of evidence is low) should always be the aim. There exist the options not to formulate a recommendation, to formulate a research recommendation, or to formulate a recommendation without grading. These are usually the exceptions.

Recommendations are strong or conditional according to the certainty of the supporting evidence and the balance between desirable and undesirable consequences of the alternative management options, the patients' values and preferences, and estimated cost or resource utilisation (Table 3, Table 4, Table 5). At KCE, typically, the KCE researchers draft the recommendation(s) which are discussed at a meeting after a presentation of the evidence review. An agreement on the direction of the effect, the strength, and the wording of the recommendation is aimed for at the meeting by reaching a consensus. If no consensus is reached at the time of the meeting, a consensus process is started via mail and followed-up at the next meeting. This process is documented in the report.

Table 3 – Strength of recommendation according to the GRADE system

Grade	Definition
Strong	The desirable effects of an intervention clearly outweigh the undesirable effects *(the intervention is to be put into practice)*, or the undesirable effects of an intervention clearly outweigh the desirable effects *(the intervention is not to be put into practice)*
Conditional	The desirable effects of an intervention probably outweigh the undesirable effects *(the intervention probably is to be put into practice)*, or the undesirable effects of an intervention probably outweigh the desirable effects *(the intervention probably is not to be put into practice)*

Source: (Andrews 2013)

Table 4 – Factors that influence the strength of a recommendation

Factor	Comment
Balance between clinical benefit and harms	The larger the difference between the desirable and undesirable effects, the higher the likelihood that a strong recommendation is warranted. The narrower the gradient, the higher the likelihood that a conditional recommendation is warranted
Quality of evidence	The higher the quality of evidence, the higher the likelihood that a strong recommendation is warranted
Values and preferences	The more values and preferences vary, or the greater the uncertainty in values and preferences, the higher the likelihood that a conditional recommendation is warranted
Costs (resource allocation)	The higher the costs of an intervention, i.e. the greater the resources consumed, the lower the likelihood that a strong recommendation is warranted

Sources: (Andrews 2013,Andrews 2013)

Table 5 – Interpretation of strong and conditional recommendations

Implications	Strong recommendation	Conditional recommendation
For patients	Most individuals in this situation would want the recommended course of action, and only a small proportion would not. Formal decision aids are not likely to be needed to help individuals make decisions consistent with their values and preferences.	The majority of individuals in this situation would want the suggested course of action, but many would not.
For clinicians	Most individuals should receive the intervention. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator.	Recognize that different choices will be appropriate for individual patients and that you must help each patient arrive at a management decision consistent with his or her values and preferences. Decision aids may be useful helping individuals making decisions consistent with their values and preferences.
For policy makers	The recommendation can be adopted as policy in most situations.	Policy-making will require substantial debate and involvement of various stakeholders.

Source: (Andrews 2013)

2.6 Good practice statements and clinical practice guidance

The GRADE Working Group developed ungraded best or good practice statement (GPS) for those situations when a guideline panel might judge the certainty high that the desirable effects of an intervention clearly outweighs its undesirable effects but the body of supportive evidence is indirect.(Dewidar 2023) In such cases, the application of the GRADE approach for grading the strength of recommendations is inappropriate.(Dewidar 2023)

The following criteria should be followed when considering a GPS: (1) the statement should be clear and actionable; (2) the message should be necessary for actual healthcare practice (ie, without the guidance, clinicians would fail to make the appropriate decision); (3) after considering all relevant outcomes and potential downstream consequences, implementing the GPS would result in large net positive consequences; (4) collection and summary of the evidence would be a poor use of a guideline panel's time and resources; and (5) a well-documented, clear and explicit rationale connecting the indirect evidence should be constructed.

Table 6 – GRADE: developing good practice statements

Process for developing good practice statement	Supporting information	Verification
Write Good Practice Statement		
Does the statement follow a PICO prioritization process?		o Yes o No
If you answered No → 1) Explain why the PICO question was not prioritized but is being addressed → avoid developing any type of actionable statement that acts as an independent response to the PICO question.		
Are the Population and Intervention components clear?		o Yes o No
If you answered No → Check whether you are developing an implementation consideration (refer to flow chart in Lotfi et al.(8))		
Items to complete	**Supporting information**	**Verification**
All the following criteria must be fulfilled *to consider the statement a GPS.*		
(1) Message is really necessary in regard to actual health care practice		This criterion has been addressed: o Yes o No
(2) Implementing the GPS results in a large net positive consequence (i.e. satisfy several evidence to decision (EtD) criteria) after consideration of all relevant outcomes and potential downstream consequences ** check the criteria considered in the development of the GPS.		This criterion has been addressed: o Yes o No **Evidence-to-Decision criteria** Importance of the problem ☐ Certainty in effects ☐ Variability and certainty in the importance of outcomes ☐ Magnitude of resource requirements ☐ Certainty of evidence of resource use ☐ Cost-effectiveness ☐ Equity ☐ Acceptability ☐ Feasibility ☐
(3) Collecting and summarizing the evidence is a poor use of a guideline panel's limited time, energy or		This criterion has been addressed: o Yes o No
resources. The opportunity cost of collecting and summarizing the evidence is large and can be avoided.		
(4) There is a well-documented clear and explicit rationale connecting the indirect evidence		This criterion has been addressed: o Yes o No
(5) Clear and actionable (This criterion should apply to any developed statement. For example, all formal recommendations are clear and actionable)		This criterion has been addressed: o Yes o No
Final judgement		
Did you answer Yes to ***all*** of the above?	o Yes Development as GPS is appropriate	
	o No Development as GPS is ***not*** appropriate → Revisit the PICO question and the importance of addressing it.	

Clinical practice guidance can add additional practical information for clinical practice identified by experts and stakeholders or by the KCE team.

2.7 External review and validation

2.7.1 Stakeholder review

At the end of the project, the results are presented to the stakeholders in a meeting or survey. A draft of the final recommendations is reviewed for clarity of the wording, feasibility and acceptability for patients, socio-cultural acceptability, feasibility for the health system, and health equity. The uptake of recommendations is also discussed.

2.7.2 Validation by external experts

An independent external review of the scientific report is performed by three experts in the research field not involved in the GDG e.g. clinical expert(s) and a methodological expert(s). This review aims to review and improve the scientific value of the report.

2.7.3 *Validation by the Belgian Centre for Evidence-Based Medicine*

The validation by the Belgian Centre for Evidence-Based Medicine (CEBAM) takes place at the end of the project after the guideline is reviewed by the external experts. The validation procedure for 'evidence-based guidelines' is based on a complete and thorough assessment of the guideline and appendices using the AGREE II instrument.

2.8 Dissemination and implementation

A range of different methods to raise awareness of the guideline exist. These include standard approaches which are applied by KCE, such as:

- issuing a press release at the moment of publication on the KCE website;
- addressing questions from the press;
- notifying experts, stakeholders and patient organisations of the publication;
- publicising the guideline through the KCE's alerts;
- using social media channels;
- and publicising the guideline's main points of interest in a journal.

For the applicability of the guideline, we refer to the section on evidence to recommendation which discusses facilitators and barriers, such as resources, cost or patient's opinions on the topic under study.

3 VALIDATED ASSESSMENT TOOLS

3.1 AGREE II – quality guidelines

The AGREE II evaluation score is used to critically appraise guidelines (Table 7). Each of the AGREE II items are rated on a 7-point scale (1–strongly disagree to 7–strongly agree). The User's Manual provides guidance on how to rate each item using the rating scale. Website AGREE and the AGREE checklist.

Table 7 – AGREE II instrument

Domain 1. Scope and Purpose
1. The overall objective(s) of the guideline is (are) specifically described.
2. The health question(s) covered by the guideline is (are) specifically described.
3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.
Domain 2. Stakeholder Involvement
4. The guideline development group includes individuals from all the relevant professional groups.
5. The views and preferences of the target population (patients, public, etc.) have been sought.
6. The target users of the guideline are clearly defined.
Domain 3. Rigour of Development
7. Systematic methods were used to search for evidence.
8. The criteria for selecting the evidence are clearly described.
9. The strengths and limitations of the body of evidence are clearly described.
10. The methods for formulating the recommendations are clearly described.
11. The health benefits, side effects, and risks have been considered in formulating the recommendations.
12. There is an explicit link between the recommendations and the supporting evidence.
13. The guideline has been externally reviewed by experts prior to its publication.
14. A procedure for updating the guideline is provided.
Domain 4. Clarity of Presentation
15. The recommendations are specific and unambiguous.
16. The different options for management of the condition or health issue are clearly presented.
17. Key recommendations are easily identifiable.
Domain 5. Applicability
18. The guideline describes facilitators and barriers to its application.
19. The guideline provides advice and/or tools on how the recommendations can be put into practice.
20. The potential resource implications of applying the recommendations have been considered.
21. The guideline presents monitoring and/ or auditing criteria.
Domain 6. Editorial Independence
22. The views of the funding body have not influenced the content of the guideline.
23. Competing interests of guideline development group members have been recorded and addressed.

3.2 AMSTAR 2 – quality systematic reviews

The AMSTAR-2 was published in 2017 by Shea et al.(Shea 2017)

Table 8 – Evaluation criteria AMSTAR-2 for a systematic review

1. Did the research questions and inclusion criteria for the review include the components of PICO?		
For Yes: ☐ Population ☐ Intervention ☐ Comparator group ☐ Outcome	Optional (recommended) ☒ Timeframe for follow-up	☐ Yes ☐ NO
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?		
For Partial Yes: The authors state that they had a written protocol or guide that included ALL the following: ☐ review question(s) ☐ a search strategy ☐ inclusion/exclusion criteria ☐ a risk of bias assessment	For Yes: As for partial yes, plus the protocol should be registered and should also have specified: ☐ a meta-analysis/synthesis plan, if appropriate, and ☐ a plan for investigating causes of heterogeneity ☐ justification for any deviations from the protocol	☐ Yes ☐ Partial Yes ☐ No
3. Did the review authors explain their selection of the study designs for inclusion in the review?		
For Yes, the review should satisfy ONE of the following: ☐ Explanation for including only RCTs ☐ OR Explanation for including only NRSI ☐ OR Explanation for including both RCTs and NRSI		☐ Yes ☐ No
4. Did the review authors use a comprehensive literature search strategy?		
For Partial Yes (all the following):	For Yes, should also have (all the following):	☐ Yes

☐ searched at least 2 databases (relevant to research question) ☐ provided key word and/or search strategy ☐ justified publication restrictions (e.g. language)	☐ searched the reference lists / bibliographies of included studies ☐ searched trial/study registries ☐ included/consulted content experts in the field ☐ where relevant, searched for grey literature ☐ conducted search within 24 months of completion of the review	☐ Partial Yes ☐ No
5. Did the review authors perform study selection in duplicate?		
For Yes, either ONE of the following: ☐ at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include ☐ OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder selected by one reviewer.		☐ Yes ☐ No
6. Did the review authors perform data extraction in duplicate?		
For Yes, either ONE of the following: ☐ at least two reviewers achieved consensus on which data to extract from included studies ☐ OR two reviewers extracted data from a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder extracted by one reviewer.		☐ Yes ☐ No
7. Did the review authors provide a list of excluded studies and justify the exclusions?		
For Partial Yes ☐ provided a list of all potentially relevant studies that were read in full-text form but excluded from the review	For Yes, must also have: ☐ Justified the exclusion from the review of each potentially relevant study	☐ Yes ☐ Partial Yes ☐ No
8. Did the review authors describe the included studies in adequate detail?		
For Partial Yes (ALL the following): ☐ described populations ☐ described interventions ☐ described comparators ☐ described outcomes	For Yes, should also have ALL the following: ☐ described population in detail ☐ described intervention in detail (including doses where relevant)	☐ Yes ☐ Partial Yes ☐ No

☐ described research design	☐ described comparator in detail (including doses where relevant) ☐ described study's setting ☐ timeframe for follow-up	
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?		
RCTs For Partial Yes, must have assessed RoB from ☐ unconcealed allocation, and ☐ lack of blinding of patients and assessors when assessing outcomes (unnecessary for objective outcomes such as all- cause mortality)	For Yes, must also have assessed RoB from: ☐ allocation sequence that was not truly random, and ☐ selection of the reported result from among multiple measurements or analyses of a specified outcome	☐ Yes ☐ Partial Yes ☐ No ☐ Includes only NRSI
NRSI For Partial Yes, must have assessed RoB: ☐ from confounding, and ☐ from selection bias	For Yes, must also have assessed RoB: ☐ methods used to ascertain exposures and outcomes, and ☐ selection of the reported result from among multiple measurements or analyses of a specified outcome	☐ Yes ☐ Partial Yes ☐ No ☐ Includes only RCTs
10. 1Did the review authors report on the sources of funding for the studies included in the review?		
For Yes ☐ Must have reported on the sources of funding for individual studies included in the review. Note: Reporting that the reviewers looked for this information. No but it was not reported by study authors also qualifies		☐ Yes ☐ No
11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?		

RCTs For Yes: ☐ The authors justified combining the data in a meta-analysis ☐ AND they used an appropriate weighted technique to combine study results and adjusted for heterogeneity if present. ☐ AND investigated the causes of any heterogeneity	☐ Yes ☐ No ☐ No meta-analysis conducted
For NRSI For Yes: ☐ The authors justified combining the data in a meta-analysis ☐ AND they used an appropriate weighted technique to combine study results, adjusting for heterogeneity if present ☐ AND they statistically combined effect estimates from NRSI that were adjusted for confounding, rather than combining raw data, or justified combining raw data when adjusted effect estimates were not available ☐ AND they reported separate summary estimates for RCTs and NRSI separately when both were included in the review	☐ Yes ☐ No ☐ No meta-analysis conducted
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	
For Yes: ☐ included only low risk of bias RCTs ☐ OR, if the pooled estimate was based on RCTs and/or NRSI at variable RoB, the authors performed analyses to investigate possible impact of RoB on summary estimates of effect.	☐ Yes ☐ No ☐ No meta-analysis conducted
13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	
For Yes: ☐ included only low risk of bias RCTs ☐ OR, if RCTs with moderate or high RoB, or NRSI were included the review provided a discussion of the likely impact of RoB on the results	☐ Yes ☐ No
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	
For Yes:	☐ Yes

☐ There was no significant heterogeneity in the results ☐ OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review	☐ No
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	
For Yes: ☐ performed graphical or statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias	☐ Yes ☐ No ☐ No meta-analysis conducted
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	
For Yes: ☐ The authors reported no competing interests OR ☐ The authors described their funding sources and how they managed potential conflicts of interest	☐ Yes ☐ No

Copied from http://amstar.ca/Amstar_Checklist.php

3.3 QUADAS-2 – risk of bias diagnostic studies

QUADAS-2 is the version of QUADAS that is recommended for use in systematic reviews to evaluate the risk of bias and applicability of primary diagnostic accuracy studies (see website).(Whiting 2011) The methodology team state on the website that in April 2025, a new version QUADAS-3 will soon be available.

QUADAS-2, consists of four key domains: patient selection, index test, reference standard, flow and timing.

Table 9 – 4 domains of the QUADAS-2 assessment

DOMAIN	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING
Description	Describe methods of patient selection: Describe included patients (prior testing, presentation, intended use of index test and setting):	Describe the index test and how it was conducted and interpreted:	Describe the reference standard and how it was conducted and interpreted:	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 interventions between index test(s) and reference standard:
Signalling questions(yes/no/unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case-control design avoided?	If a threshold was used, was it pre-specified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
Risk of bias: High/low/unclear	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability: High/low/unclear	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

Table copied from Bristol University.

Table 10 – Quality of diagnostic studies with QUADAS-2 assessment

Signalling Questions	Answers
Participant selection - allocation	
1) Was a consecutive or random sample of participants or images enrolled?	Yes - if paper states consecutive or random; No – if paper describes other method of sampling; Unclear - if not described
2) Was a case-control design avoided?	Yes - if consecutive or random or case-control design clearly not used; No – if study described as case-control or describes sampling specific numbers of participants with particular diagnoses; Unclear - if not described
3) Did the study avoid inappropriate exclusions?	Yes - if inappropriate exclusions were avoided; No – if lesions were excluded that might affect test accuracy, e.g. indeterminate results or where disagreement between evaluators was observed; Unclear – if not clearly reported
Could the selection of participants have introduced bias?	*If answers to all of questions 1) and 2) and 3) was 'Yes': Risk is Low* *If answers to any one of questions 1) or 2) or 3) was 'No': Risk is High* *If answers to any one of questions 1) or 2) or 3) was 'Unclear': Risk Unclear*
Participant selection - applicability	
Is there concern that the included participants do not match the review question?	*If the answer to question is 'Yes': Concern is High* *If the answer to question is 'No': Concern is Low* *If the answer to question is 'Unclear': Concern is Unclear*
Index test – risk of bias	
1) Was the index test or testing strategy result interpreted without knowledge of the results of the reference standard?	Yes - if index test described as interpreted without knowledge of reference standard result, or for prospective studies, if index test is always conducted and interpreted before the reference standard; No – if index test described as interpreted in knowledge of reference standard result; Unclear – if index test blinding is not described
2) Was the diagnostic threshold at which the test was considered positive prespecified?	Yes - if threshold was prespecified (i.e. before analysing study results); No - if threshold was not prespecified; Unclear - if not possible to tell whether or not diagnostic threshold was prespecified
Could the conduct or interpretation of the index test have introduced bias?	*If answers to questions 1) and 2) was 'Yes': Risk is Low* *If answers to either questions 1) or 2) was 'No': Risk is High* *If answers to either questions 1) or 2) was 'Unclear': Risk is Unclear*
Index test – applicability	
1) Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? This item applies equally to studies using objective and more subjective approaches to test interpretation.	Yes – if the criteria for diagnosis of the target disorder were reported in sufficient detail to allow replication No – if the criteria for diagnosis of the target disorder were not report- ed in sufficient detail to allow replication Unclear – if some but not sufficient information on criteria for diagnosis to allow replication were provided
2) Was the test interpretation carried out by an experienced examiner?	Yes – if the test was interpreted by an experienced examiner as mentioned in paper No – if the test was not interpreted by an experienced examiner Unclear – if the experience of the examiner(s) was not reported in sufficient detail to judge or if examiners described as 'Expert' with no further detail given
Is there concern that the index test, its conduct, or interpretation differ from the review question?	*If answers to questions 1) and 2) was 'Yes': Concern is Low* *If answers to questions 1) or 2) was 'No': Concern is High* *If answers to questions 1) or 2) was 'Unclear': Concern is Unclear*
Reference standard – risk of bias	
1) Is the reference standard likely to correctly classify the target condition?	
2) Were the reference standard results interpreted without knowledge of the results of the index test?	Yes – if described as blinded to the index test result No – if described as having knowledge of the index test result Unclear – if interpretation was not clearly reported
Could the reference standard, its conduct, or its interpretation have introduced bias?	*If answers to questions was 'Yes': Risk is Low* *If answers to questions was 'No': Risk is High* *If answers to questions was 'Unclear': Risk is Unclear*

Reference standard - applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	*If answer is 'Yes': Concern is High* *If answer is 'No': Concern is Low* *If answer is 'Unclear': Concern is Unclear*
Flow and timing - risk of bias	
1) Was there an appropriate interval between index test and reference standard?	
2) Did all participants receive the same reference standard?	Yes – if all participants underwent the same reference standard; No – if some participants were excluded from the analysis Unclear – if not clearly reported
3) Were all participants included in the analysis?	Yes – if all participants were included in the analysis No – if some participants were excluded from the analysis Unclear – if not clearly reported
Could the participant flow have introduced bias?	*If answers to questions 1) and 2) and 3) was 'Yes': Risk is Low* *If answers to any one of questions 1) or 2) or 3) was 'No': Risk is High* *If answers to any one of questions 1) or 2) or 3) was 'Unclear': Risk is Unclear*

3.4 ROB-1 – Risk of bias for RCTs

Risk of bias (RoB-1) can be used as the basis for the evaluation.(Higgins, Altman, on behalf of the Cochrane Statistical Methods Group and the Cochrane Bias Methods Group 2012)

Table 11 – Cochrane Collaboration's risk of bias tool 1 for assessing risk of bias

Domain	Support for judgement	Review authors' judgement
Selection bias		
Random sequence generation	Describe the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups	Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence
Allocation concealment	Describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen in advance of, or during, enrolment	Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment
Performance bias		
Blinding of participants and personnel Assessments should be made for each main outcome (or class of outcomes)	Describe all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective	Performance bias due to knowledge of the allocated interventions by participants and personnel during the study
Detection bias		
Blinding of outcome assessment Assessments should be made for each main outcome (or class of outcomes)	Describe all measures used, if any, to blind outcome assessors from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective	Detection bias due to knowledge of the allocated interventions by outcome assessors
Attrition bias		
Incomplete outcome data Assessments should be made for each main outcome (or class of outcomes)	Describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomised participants), reasons for attrition/exclusions where reported, and any reinclusions in analyses performed by the review authors	Attrition bias due to amount, nature or handling of incomplete outcome data
Reporting bias		
Selective reporting	State how the possibility of selective outcome reporting was examined by the review authors, and what was found	Reporting bias due to selective outcome reporting
Other bias		
Other sources of bias	State any important concerns about bias not addressed in the other domains in the tool If particular questions/entries were prespecified in the review's protocol, responses should be provided for each question/entry	Bias due to problems not covered elsewhere in the table

3.5 ROBINS-I – Risk Of Bias In Non-randomised Studies of Interventions

Per outcome, seven domains need to be assessed: confounding bias, selection bias, bias in classification of interventions, bias due to deviations from intended intervention, bias due to missing data, bias in measurement outcomes and bias in selection of the reported result. A key feature of the tool is the inclusion of signalling questions within each domain of bias. The overall risk of bias judgement of a study outcome is based on the response options for each domain level. The risk of bias judgements are: low risk of bias (the study is comparable to a well-performed randomised controlled trial), moderate risk of bias (the study is sound for a non-randomised controlled trial but cannot be compared to a RCT), serious risk of bias (the study has some important problems), critical risk of bias (the study is too problematic to provide any useful evidence on the effects of the intervention) or no information to base a judgement. When at least one individual domain is considered at a serious risk of bias, the overall risk will become serious risk of bias, irrespective of other more 'positively judged' domains. In general, it rarely happens that a non-randomised controlled trial is judged as having a low risk of bias.

REFERENCES

1. AGREE Next Steps Consortium 2017. The AGREE II Instrument [Electronic version]. Retrieved September 24th, 2024 from: http://www.agreetrust.org.
2. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *Journal of clinical epidemiology.* Jul 2013;66(7):719-725.
3. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *Journal of clinical epidemiology.* Jul 2013;66(7):726-735.
4. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* Dec 14 2010;182(18):E839-842.
5. Dewidar O, Lotfi T, Langendam MW, et al. Good or best practice statements: proposal for the operationalisation and implementation of GRADE guidance. *BMJ Evid Based Med.* Jun 2023;28(3):189-196.
6. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. *Journal of clinical epidemiology.* Dec 2011;64(12):1283-1293.
7. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *Journal of clinical epidemiology.* Dec 2011;64(12):1303-1310.
8. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *Journal of clinical epidemiology.* Dec 2011;64(12):1294-1302.
9. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. *Journal of clinical epidemiology.* Dec 2011;64(12):1277-1282.
10. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. *Journal of clinical epidemiology.* 2011/04/01/ 2011;64(4):380-382.
11. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *Journal of clinical epidemiology.* Apr 2011;64(4):407-415.
12. Higgins JPT, Altman DG, on behalf of the Cochrane Statistical Methods Group and the Cochrane Bias Methods Group. *Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors*: Cochrane handbook for systematic reviews of interventions: John Wiley & Sons; 2012. p. 187-241; 2012.
13. Klugar M, Lotfi T, Darzi AJ, et al. GRADE guidance 39: using GRADE-ADOLOPMENT to adopt, adapt or create contextualized recommendations from source guidelines and evidence syntheses. *Journal of clinical epidemiology.* Oct 2024;174:111494.
14. Piggott T, Baldeh T, Dietl B, et al. Standardized wording to improve efficiency and clarity of GRADE EtD frameworks in health guidelines. *Journal of clinical epidemiology.* 2022/06/01/ 2022;146:106-122.
15. Schünemann HJ, Wiercioch W, Brozek J, et al. GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT. *Journal of clinical epidemiology.* Jan 2017;81:101-110.
16. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical research ed.).* Sep 21 2017;358:j4008.
17. Skoetz N, Goldkuhle M, van Dalen EC, et al. GRADE guidelines 27: how to calculate absolute effects for time-to-event outcomes in summary of findings tables and Evidence Profiles. *Journal of clinical epidemiology.* Feb 2020;118:124-131.
18. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* Oct 12 2016;355:i4919.
19. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine.* Oct 18 2011;155(8):529-536.