

Colorectal cancer (update)

Supplement 2: Methods

NICE guideline TBC

Methods

July 2019

Draft for Consultation

Developed by the National Guideline Alliance hosted by the Royal College of Obstetricians and Gynaecologists

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the [Welsh Government](#), [Scottish Government](#), and [Northern Ireland Executive](#). All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

© NICE 2019. All rights reserved. Subject to [Notice of rights](#).

ISBN:

Contents

Development of the guideline.....	5
Remit.....	5
Declarations of interest.....	5
What this guideline covers.....	5
Groups that are covered.....	5
Clinical areas that are covered.....	5
What this guideline does not cover.....	6
Groups that are not covered.....	6
Clinical areas that are not covered.....	6
Methods.....	8
Developing the review questions and outcomes.....	8
Searching for evidence.....	10
Clinical literature search.....	10
Health economic literature search.....	12
Reviewing research evidence.....	12
Systematic review process.....	12
Type of studies and inclusion/exclusion criteria.....	13
Methods of combining evidence.....	13
Appraising the quality of evidence.....	15
Reviewing economic evidence.....	24
Inclusion and exclusion of economic studies.....	24
Appraising the quality of economic evidence.....	24
Health economic modelling.....	24
Expert evidence.....	25
Developing recommendations.....	26
Guideline recommendations.....	26
Research recommendations.....	26
Validation process.....	26
Updating the guideline.....	26
Funding.....	26
References.....	27

1 Development of the guideline

2 Remit

3 The National Institute for Health and Care Excellence (NICE) commissioned the
4 National Guideline Alliance (NGA) to develop a guideline to update and replace the
5 NICE guideline on colorectal cancer: diagnosis and management (CG131) and the
6 NICE guideline on improving outcomes in colorectal cancer (CSG5).

7 Declarations of interest

8 Committee members' and developers' declarations of interest were recorded
9 according to NICE's 2014 conflicts of interest policy until 31st March 2018, and
10 thereafter in accordance with NICE's 2018 conflicts of interest policy.

11 What this guideline covers

12 Groups that are covered

- 13 • Adults (18 years and older) with newly diagnosed adenocarcinoma of the colon
- 14 • Adults with newly diagnosed adenocarcinoma of the rectum
- 15 • Adults with relapsed adenocarcinoma of the colon
- 16 • Adults with relapsed adenocarcinoma of the rectum
- 17 • Adults with clinical or genetic evidence of Lynch syndrome (hereditary
18 nonpolyposis colorectal cancer, HNPCC)

19 Clinical areas that are covered

20 The guideline covers the following clinical issues:

- 21 • Prevention of colorectal cancer
 - 22 ○ Role of aspirin in the prevention of colorectal cancer in adults with clinical or
23 genetic evidence of Lynch syndrome (hereditary nonpolyposis colorectal
24 cancer)
- 25 • Molecular biomarkers
 - 26 ○ Use of molecular biomarkers to guide chemotherapy choice
- 27 • Management of local disease
 - 28 ○ Rectal cancer
 - 29 ○ Colon cancer
 - 30 ○ Colonic stents for obstructing colon cancer
- 31 • Management of metastatic disease
 - 32 ○ Presenting with stage IV colorectal cancer
 - 33 ○ Methods for treating metastasis
- 34 • Ongoing care and support
 - 35 ○ Follow-up after apparently curative resection
 - 36 ○ Management of post treatment sequelae
 - 37 ○ Information about managing bowel function
- 38 • Service delivery

- 1 ○ Surgical volumes and rectal cancer surgery
- 2 For further details please refer to the [scope](#) on the NICE website.

3 **What this guideline does not cover**

4 **Groups that are not covered**

5 The guideline does not cover the following groups:

- 6 • People with anal cancer
- 7 • Children and young people aged under 18 years
- 8 • People with primary or secondary lymphoma of the colon and rectum
- 9 • People with pure small cell carcinoma, or other pure neuroendocrine carcinomas,
10 of the colon and rectum
- 11 • People with neuroendocrine tumours of the colon and rectum
- 12 • People with gastrointestinal stromal tumours (GIST) or sarcoma of the colon and
13 rectum
- 14 • People with squamous cells carcinoma of the rectum
- 15 • People with appendiceal neoplasms

16 **Clinical areas that are not covered**

17 This guideline does not cover the following areas:

- 18 • Population screening
- 19 • Colonoscopic surveillance of high-risk groups, including people with a family
20 history of colorectal cancer and people with inflammatory bowel disease
- 21 • Management of anal cancer

22 The following areas covered by CG131 were not updated and will be removed from
23 the guideline as there is no longer variation in practice:

- 24 • Diagnostic investigations
- 25 • Staging of colorectal cancer
- 26 • Imaging of hepatic metastases
- 27 • Imaging of extra-hepatic metastases

28 The following areas from CSG5 will not be updated either because they are already
29 covered within scope of update of CG131 or other NICE guidelines or because they
30 are no longer relevant to this guideline:

- 31 • Patient centred care
- 32 • Access to appropriate services
- 33 • Multidisciplinary teams
- 34 • Diagnosis
- 35 • Surgery and histopathology
- 36 • Radiotherapy in primary disease
- 37 • Adjuvant chemotherapy
- 38 • Anal cancer
- 39 • Follow-up

- 1 • Recurrent and advanced disease
- 2 • Palliative care
- 3 For further details please refer to the [scope](#) on the NICE website.

1 Methods

2 Introduction

3 This section summarises methods used to identify and review the evidence, to
4 consider cost effectiveness, and to develop guideline recommendations. This
5 guideline was developed in accordance with methods described in [Developing NICE](#)
6 [guidelines: the manual](#) (NICE 2014).

7 Developing the review questions and outcomes

8 The review questions considered in this guideline were based on the key areas
9 identified in the guideline [scope](#). They were drafted by the NGA technical team, and
10 refined and validated by the guideline committee.

11

12 The review questions were based on the following frameworks:

- 13 • intervention reviews – using population, intervention, comparison and outcome
14 (PICO)
- 15 • prognostic reviews – using population, presence or absence of a prognostic, risk
16 or predictive factor and outcome (PPO)
- 17 • qualitative review – using population, phenomenon of interest and context (PICo).

18 These frameworks guided the development of review protocols, the literature
19 searching process, and critical appraisal and synthesis of evidence. They also
20 facilitated development of recommendations by the committee.

21 Full literature searches, critical appraisal and evidence reviews were completed for
22 all review questions.

23 The review questions and evidence review reports corresponding to each question
24 (or group of questions) are summarised in Table 1.

25 **Table 1: Summary of review questions and index to evidence reports**

Evidence report	Review question	Type of review
A. Prevention of colorectal cancer		
A1. Effectiveness of aspirin in the prevention of colorectal cancer in people with Lynch syndrome	How effective is aspirin in the prevention of colorectal cancer in adults with Lynch syndrome (hereditary nonpolyposis colorectal cancer)?	Intervention
B. Molecular biomarkers		
B1. Use of molecular biomarkers to guide systemic therapy	Which predictive biomarkers should be used in the systemic management of colorectal cancer patients?	Predictive/prognostic
C. Management of local disease		
C1. Treatment for early rectal cancer	What is the most effective treatment for early rectal cancer?	Intervention
C2. Preoperative	What is the effectiveness of preoperative	Intervention

Evidence report	Review question	Type of review
radiotherapy and chemoradiotherapy for rectal cancer	radiotherapy or chemo radiotherapy for rectal cancer?	
C3. Optimal surgical technique for rectal cancer	What is the optimal surgical technique for rectal cancer?	Intervention
C4. Deferral of surgery in people having neoadjuvant therapy for rectal cancer	Which people having neoadjuvant chemotherapy or chemoradiotherapy for rectal cancer do not need surgery?	Prognostic
C5. Effectiveness of exenterative surgery for locally advanced or recurrent rectal cancer	What is the effectiveness of exenterative surgery for locally advanced or recurrent rectal cancer?	Intervention
C6. Endoscopic resection alone for early colon cancer	Which people with early colon cancer can be treated with endoscopic resection alone?	Intervention
C7. Preoperative chemotherapy for non-metastatic colon cancer	Which people with non-metastatic colon cancer would benefit from preoperative chemotherapy?	Intervention
C8. Optimal duration of adjuvant chemotherapy for colorectal cancer	What is the optimal duration of adjuvant chemotherapy for colorectal cancer?	Intervention
C9. Effectiveness of stenting for acute large bowel obstruction	What is the effectiveness of stenting compared with emergency surgery for suspected colorectal cancer causing acute large bowel obstruction?	Intervention
D. Management of metastatic disease		
D1. Surgery for asymptomatic primary tumour	Does surgery for the asymptomatic primary tumour improve outcomes for people with metastatic colorectal cancer, which cannot be treated with curative intent?	Intervention
D2a. Treatment for metastatic colorectal cancer in the liver amenable to treatment with curative intent	What is the optimal combination and sequence of treatments in patients presenting with metastatic colorectal cancer in the liver amenable to treatment with curative intent?	Intervention
D2b. Treatment for metastatic colorectal cancer in the liver not amenable to treatment with curative intent	What is the optimal combination and sequence of treatments in patients presenting with metastatic colorectal cancer in the liver not amenable to treatment with curative intent?	Intervention
D3. Treatment for	What is the optimal combination and sequence	Intervention

Evidence report	Review question	Type of review
metastatic colorectal cancer in the lung amenable to local treatment	of treatments in patients presenting with metastatic colorectal cancer in the lung amenable to local treatment?	
D4. Local and systemic treatments for metastatic colorectal cancer isolated in the peritoneum	What is the optimal combination and sequence of local and systemic treatments in patients presenting with metastatic colorectal cancer isolated in the peritoneum?	Intervention
E. Ongoing care and support		
E1. Follow-up to detect recurrence after treatment for non-metastatic colorectal cancer	What are the optimal methods and frequencies of follow-up to detect recurrence after potentially curative surgical treatment for non-metastatic colorectal cancer?	Intervention
E2. Optimal management of low anterior resection syndrome	What is the optimal management of low anterior resection syndrome?	Intervention
E3. Information needs of people prior, during and after treatment for colorectal cancer	What are the information needs of people prior, during and after treatment for colorectal cancer?	Qualitative
F. Service delivery		
F1. Surgical volumes and outcomes for rectal cancer	Is there a relationship between surgical volumes and outcomes in the treatment of rectal cancer (primary and recurrent disease)?	Predictive/prognostic

1 Additional information related to development of the guideline is contained in:

- 2 • Supplement 1: Methods (this document)
- 3 • Supplement 2: Health economics
- 4 • Supplement 3: Glossary
- 5 • Supplement 4: NGA technical team list.

6 Searching for evidence

7 Clinical literature search

8 Systematic literature searches were undertaken to identify published clinical
9 evidence relevant to each review question. Combined searches were more than one
10 review question were conducted where appropriate. A combined search was done for
11 evidence reviews C1, C2 and C3; and for evidence reviews D2a and D2b.

12 Databases were searched using medical subject headings, free-text terms and study
13 type filters where appropriate. Where possible, searches were restricted to retrieve
14 articles published in English. All searches were conducted in the following databases:

1 Medline, Medline-in-Process, Cochrane Central Register of Controlled Trials (CCTR),
2 Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of
3 Reviews of Effects (DARE), Health Technology Assessments (HTA) and Embase.
4 For review questions related to information provision, PsycInfo, CINAHL and Web of
5 Science were also searched. Web of Science was also used for the question about
6 prevention of colorectal cancer.

7 Searches were run once for all reviews during development. Searches were updated
8 6 to 8 weeks in advance of the final committee meetings for the following questions:

- 9 • C1. What is the most effective treatment for early rectal cancer?
- 10 • C2. What is the effectiveness of preoperative radiotherapy or chemo radiotherapy
11 for rectal cancer?
- 12 • C3. What is the optimal surgical technique for rectal cancer?
- 13 • C4. Which people having neoadjuvant chemotherapy or chemoradiotherapy for
14 rectal cancer do not need surgery?
- 15 • C5. What is the effectiveness of exenterative surgery for locally advanced or
16 recurrent rectal cancer?
- 17 • C8. What is the optimal duration of adjuvant chemotherapy for colorectal cancer?
- 18 • D1. Does surgery for the asymptomatic primary tumour improve outcomes for
19 people with metastatic colorectal cancer, which cannot be treated with curative
20 intent?
- 21 • D2a. What is the optimal combination and sequence of treatments in patients
22 presenting with metastatic colorectal cancer in the liver amenable to treatment
23 with curative intent?
- 24 • D2b. What is the optimal combination and sequence of treatments in patients
25 presenting with metastatic colorectal cancer in the liver not amenable to treatment
26 with curative intent?
- 27 • E1. What are the optimal methods and frequencies of follow-up to detect
28 recurrence after potentially curative surgical treatment for non-metastatic
29 colorectal cancer?
- 30 • E2. What is the optimal management of low anterior resection syndrome?
- 31 • F1. Is there a relationship between surgical volumes and outcomes in the
32 treatment of rectal cancer (primary and recurrent disease)?

33 Literature searches were not updated for the remaining review questions because:

- 34 • the original search was done within 8 weeks of the final committee meeting
- 35 • there was already robust evidence and new evidence would not change the
36 conclusions or the committee was confident there was no new published
37 evidence.

38 Search strategies were quality assured by cross-checking reference lists of relevant
39 articles, analysing search strategies from other systematic reviews and asking
40 members of the committee to highlight key studies. All search strategies were also
41 quality assured by an information scientist who was not involved in developing the
42 primary search strategy. Details of the search strategies, including study-design
43 filters applied and databases searched, are presented in Appendix B of each
44 evidence report.

45 All publications highlighted by stakeholders at the time of the consultation on the draft
46 scope were considered for inclusion. During the scoping phase, searches were
47 conducted for guidelines, health technology assessments, systematic reviews,

1 economic evaluations and reports on biomedical databases and websites of
2 organisations relevant to the topic. Formal searching for grey literature and
3 unpublished literature was not undertaken routinely.

4 **Health economic literature search**

5 A global search of economic evidence was undertaken and re-run in May 2019. The
6 following databases were searched:

- 7 • MEDLINE (Ovid)
- 8 • EMBASE (Ovid)
- 9 • Health Technology Assessment database (HTA)
- 10 • NHS Economic Evaluation Database (NHS EED)

11 Further to the database searches, the committee was contacted with a request for
12 details of relevant published and unpublished studies of which they may have
13 knowledge; reference lists of key identified studies were also reviewed for any
14 potentially relevant studies. Finally, the NICE website was searched for any recently
15 published guidance relating to colorectal cancer that had not been already identified
16 via the database searches.

17 The search strategy for existing economic evaluations combined terms capturing
18 colorectal cancer and, for searches undertaken in MEDLINE and EMBASE, terms to
19 capture economic evaluations. No restrictions on language or setting were applied to
20 the economic evidence search, but a standard exclusions filter was applied (letters,
21 animals, etc.). Full details of the search strategy are presented in Supplement 2:
22 Health economics.

23 **Reviewing research evidence**

24 **Systematic review process**

25 The evidence was reviewed in accordance with the following approach.

- 26 • Potentially relevant articles were identified from the search results for each review
27 question by screening titles and abstracts. Full-text copies of the articles were
28 then obtained.
- 29 • Full-text articles were reviewed against pre-specified inclusion and exclusion
30 criteria in the review protocol (see appendix A of each evidence report).
- 31 • Key information was extracted from each article on study methods and results, in
32 accordance with factors specified in the review protocol. The information was
33 presented in a summary table in the corresponding evidence report and in a more
34 detailed evidence table (see appendix D of each evidence report).
- 35 • Included studies were critically appraised using an appropriate checklist as
36 specified in [Developing NICE guidelines: the manual](#) (NICE 2014). Further detail
37 on appraisal of the evidence is provided below.
- 38 • Summaries of evidence by outcome were presented in the corresponding
39 evidence report and discussed by the committee.

40 All review questions were subject to dual screening and study selection through a
41 10% random sample of articles, except for the question on effectiveness of aspirin in the
42 prevention of colorectal cancer in adults with Lynch syndrome. This question was not
43 dual screened because the intervention, population and study design were well

1 defined. Any discrepancies in dual screening were resolved by discussion between
2 the first and second reviewers or by reference to a third (senior) reviewer. Internal
3 (NGA) quality assurance processes included consideration of the outcomes of
4 screening, study selection and data extraction and the committee reviewed the
5 results of study selection and data extraction. The review protocol for each question
6 specifies whether dual screening and study selection was undertaken for that
7 particular question.

8 Drafts of all evidence reviews were checked by a senior reviewer.

9 **Type of studies and inclusion/exclusion criteria**

10 Inclusion and exclusion of studies was based on criteria specified in the
11 corresponding review protocol.

12 Systematic reviews with meta-analyses were considered to be the highest quality
13 evidence that could be selected for inclusion.

14 For intervention reviews, randomised controlled trials (RCTs) were prioritised for
15 inclusion because they are considered to be the most robust type of study design
16 that could produce an unbiased estimate of intervention effects. Where there was
17 limited evidence from RCTs, non-randomised controlled trials were considered for
18 inclusion.

19 For prognostic reviews, prospective and retrospective cohort and case–control
20 studies and case series were considered for inclusion.

21 For the qualitative review, studies using focus groups, structured interviews or semi-
22 structured interviews were considered for inclusion. Where qualitative evidence was
23 sought, data from surveys or other types of questionnaire were considered for
24 inclusion only if they provided data from open-ended questions, but not if they
25 reported only quantitative data.

26 The committee was consulted about any uncertainty regarding inclusion or exclusion
27 of studies. A list of excluded studies for each review question, including reasons for
28 exclusion is presented in appendix K of the corresponding evidence report.

29 Narrative reviews, posters, letters, editorials, comment articles, unpublished studies
30 and studies published in languages other than English were excluded. Conference
31 abstracts were not considered for inclusion.

32 **Methods of combining evidence**

33 When planning reviews (through preparation of protocols), the following approaches
34 for data synthesis were discussed and agreed with the committee.

35 **Data synthesis for intervention studies**

36 Meta-analysis to pool results from RCTs was conducted where possible using
37 Cochrane Review Manager (RevMan5) software. Where non-randomised evidence
38 was used, this was meta-analysed only if results had been adjusted for pre-specified
39 confounders.

40 For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a
41 fixed effect model was used to calculate risk ratios (relative risks; RRs). For all
42 outcomes with zero events in both arms the risk difference was presented. For

1 outcomes with zero events in only one arm, Peto odds ratios (ORs) were calculated
2 as this method performs well when events are rare (Bradburn 2007).

3 In some cases outcomes were meta-analysed for a single study arm only. This was
4 when the outcome was only possible in one study arm, for example the technical
5 success of stenting in trials comparing stents to palliative care. Studies were only
6 meta-analysed if they had similar populations, interventions and outcome definitions.
7 Proportions were first logit transformed and then meta-analysed with a random-
8 effects model using the metafor package in R (Viechtbauer 2010).

9 Time-to-event outcomes, such as overall survival, were meta-analysed using the “O
10 – E (Observed – Expected) and Variance” outcome type in RevMan5. This provides
11 a fixed-effect meta-analysis of log hazard ratios. Corrections for zero cell counts are
12 not needed when using this method. If O – E and Variance were not reported in a
13 study, where possible these were calculated using methods reported by Tierney
14 (2007). In some cases where studies had used Cox proportional hazards regression
15 models, generic inverse variance meta-analysis was used. Both fixed-effect and
16 random-effects analyses were available in this case.

17 For continuous outcomes, measures of central tendency (mean) and variation
18 (standard deviation; SD) are required for meta-analysis. Data for continuous
19 outcomes, such as duration of hospital stay, were meta-analysed using an inverse-
20 variance method for pooling weighted mean differences (WMDs). Where SDs were
21 not reported for each intervention group, the standard error (SE) of the mean
22 difference was calculated from other reported statistics (p values or 95% confidence
23 intervals; CIs) and then meta-analysis was conducted as described above.

24 When evidence was based on studies that reported descriptive data or medians with
25 interquartile ranges or p values, this information was included in the corresponding
26 GRADE tables (see below) without calculating relative or absolute effects.
27 Consequently, certain aspects of quality assessment such as imprecision of the
28 effect estimate could not be assessed for this type of evidence. The limited reporting
29 was interpreted as representing a risk of bias when assessing study limitations.

30 Subgroups for stratified analyses were agreed for some review questions as part of
31 protocol development.

32 When meta-analysis was undertaken, the results were presented visually using forest
33 plots generated using RevMan5 (see appendix E of relevant evidence reports).

34 **Data synthesis for prognostic reviews**

35 Odds ratios (ORs) or RRs with 95% CIs reported in published studies were extracted
36 or calculated by the NGA technical team to examine relationships between risk
37 factors and outcomes of interest. Ideally analyses would have adjusted for key
38 confounders to be considered for inclusion. Prognostic data were only pooled if
39 studies were similar in terms of populations, risk factors, outcomes and statistical
40 analysis methods (including adjustments for confounding factors), otherwise results
41 from individual studies were presented in the evidence reports.

42 **Data synthesis for qualitative reviews**

43 Where possible, a meta-synthesis was conducted to combine evidence from
44 qualitative studies. Whenever studies identified a qualitative theme relevant to the
45 protocol, this was extracted and the main characteristics were summarised. When all
46 themes had been extracted from studies, common concepts were categorised and

1 tabulated. This included information on how many studies had contributed to each
2 theme identified by the NGA technical team.

3 In qualitative synthesis, a theme being reported more than other themes across
4 included studies does not necessarily mean that the theme is more important than
5 other themes. The aim of qualitative research is to identify new perspectives on a
6 particular topic. Study types and populations in qualitative research can differ widely,
7 meaning that themes identified by just one or a few studies can provide important
8 new information on a given topic.

9 Themes from individual studies were integrated into a wider context and, when
10 possible, overarching categories of themes with sub-themes were identified. Themes
11 were derived from data presented in individual studies. When themes were extracted
12 from 1 primary study only, theme names used in the guideline mirrored those in the
13 source study. However, when themes were based on evidence from multiple studies,
14 the theme names were assigned by the NGA technical team. The names of
15 overarching categories of themes were also assigned by the NGA technical team.

16 Emerging themes were placed into a thematic map representing the relationship
17 between themes and overarching categories. The purpose of such a map is to show
18 relationships between overarching categories and associated themes.

19 **Appraising the quality of evidence**

20 **Intervention studies**

21 ***GRADE methodology for intervention reviews***

22 For intervention reviews, the evidence for outcomes from included RCTs and
23 comparative non-randomised studies was evaluated and presented using the
24 Grading of Recommendations Assessment, Development and Evaluation (GRADE)
25 methodology developed by the international [GRADE working group](#).

26 When GRADE was applied, software developed by the GRADE working group
27 (GRADEpro) was used to assess the quality of each outcome, taking account of
28 individual study quality factors and any meta-analysis results. Results were
29 presented in GRADE profiles (GRADE tables).

30 The selection of outcomes for each review question was agreed during development
31 of the associated review protocol in discussion with the committee. The evidence for
32 each outcome was examined separately for the quality elements summarised in
33 Table 2. Criteria considered in the rating of these elements are discussed below.
34 Each element was graded using the quality ratings summarised in Table 3. Footnotes
35 to GRADE tables were used to record reasons for grading a particular quality
36 element as having a 'serious' or 'very serious' quality issue. The ratings for each
37 component were combined to obtain an overall assessment of quality for each
38 outcome as described in Table 4.

39 The initial quality rating was based on the study design: RCTs start as 'high' quality
40 evidence and non-randomised studies as 'low' quality evidence. The rating was then
41 modified according to the assessment of each quality element (Table 2). Each quality
42 element considered to have a 'serious' or 'very serious' quality issue was
43 downgraded by 1 or 2 levels respectively (for example, evidence starting as 'high'
44 quality was downgraded to 'moderate' or 'low' quality). In addition, there was a
45 possibility to upgrade evidence from non-randomised studies (provided the evidence
46 for that outcome had not previously been downgraded) if there was a large

1 magnitude of effect, a dose–response gradient, or if all plausible confounding would
 2 reduce a demonstrated effect or suggest a spurious effect when results showed no
 3 effect.

4 **Table 2: Summary of quality elements in GRADE for intervention reviews**

Quality element	Description
Risk of bias ('Study limitations')	Limitations in study design and implementation may bias estimates of treatment effect. High risk of bias for the majority of the evidence reduces confidence in the estimated effect
Inconsistency	This refers to unexplained heterogeneity in the results
Indirectness	This refers to differences in study populations, interventions, comparators or outcomes between the available evidence and inclusion criteria specified in the review protocol
Imprecision	This occurs when a study has relatively few participants or few events of interest, resulting in wide confidence intervals around estimates of effect that include clinically important thresholds
Publication bias	This refers to systematic under- or over-estimation of the underlying benefit or harm resulting from selective publication of study results

5 **Table 3: GRADE quality ratings (by quality element)**

Quality issues	Description
None or not serious	No serious issues with the evidence for the quality element under consideration
Serious	Issues with the evidence sufficient to downgrade by 1 level for the quality element under consideration
Very serious	Issues with the evidence sufficient to downgrade by 2 levels for the quality element under consideration

6 **Table 4: Overall quality of the evidence in GRADE (by outcome)**

Overall quality grading	Description
High	Further research is very unlikely to change the level of confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on the level of confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on the level of confidence in the estimate of effect and is likely to change the estimate
Very low	The estimate of effect is very uncertain

7 *Assessing risk of bias in intervention reviews*

8 Bias is a systematic error, or consistent deviation from the truth in results obtained.
 9 When a risk of bias is present the true effect can be either under- or over-estimated.

10 Risk of bias in RCTs was assessed using the Cochrane risk of bias tool (see
 11 Appendix H in [Developing NICE guidelines: the manual](#); NICE 2014).

12 The Cochrane risk of bias tool assesses the following possible sources of bias:

- 1 • selection bias
- 2 • performance bias
- 3 • attrition bias
- 4 • detection bias
- 5 • reporting bias.

6 A study with a poor methodological design does not automatically imply high risk of
7 bias; the bias is considered individually for each outcome and it is assessed whether
8 the chosen design and methodology will impact on the estimation of the intervention
9 effect. For example for outcomes considered objective (survival, complete resection
10 rate, local recurrence, sphincter preservation/permanent stoma and treatment-related
11 mortality) we did not automatically downgrade for lack of blinding.

12 More details about the Cochrane risk of bias tool can be found in Section 8 of the
13 [Cochrane Handbook for Systematic Reviews of Interventions](#) (Higgins 2011).

14 For systematic reviews the ROBIS checklist was used (see Appendix H in
15 [Developing NICE guidelines: the manual](#); NICE 2014).

16 For non-randomised studies the ROBINS-I checklist was used (see Appendix H in
17 [Developing NICE guidelines: the manual](#); NICE 2014).

18 When meta-analysis was performed the respective weights of the individual studies
19 in the meta-analysis were considered when assessing the quality of the evidence.
20 For example, when an individual study had a serious risk of bias but only accounted
21 for, for example, 2% of the weight of the meta-analysis, the evidence was not
22 downgraded for serious risk of bias.

23 *Assessing inconsistency in intervention reviews*

24 Inconsistency refers to unexplained heterogeneity in results of meta-analysis. When
25 estimates of treatment effect vary widely across studies (that is, there is
26 heterogeneity or variability in results), this suggests true differences in underlying
27 effects. Inconsistency is, thus, only truly applicable when statistical meta-analysis is
28 conducted (that is, results from different studies are pooled). When outcomes were
29 derived from a single study the rating 'no serious inconsistency' was used when
30 assessing this domain, as per GRADE methodology (Santesso 2016).

31 Inconsistency was assessed visually by inspecting forest plots and observing
32 whether there was considerable heterogeneity in the results of the meta-analysis.
33 This was assessed by calculating the I-squared statistic for the meta-analysis with an
34 I-squared value of more than 50% indicating considerable heterogeneity, and more
35 than 80% indicating very serious heterogeneity. When considerable or very serious
36 heterogeneity was observed, possible reasons were explored and subgroup analyses
37 were performed as pre-specified in the review protocol where possible. In the case of
38 unexplained heterogeneity, sensitivity analyses were planned based on the quality of
39 studies, eliminating studies at high risk of bias (in relation to randomisation, allocation
40 concealment and blinding, and/or missing outcome data).

41 When considerable heterogeneity was present, the meta-analysis was re-run using
42 the Der-Simonian and Laird method with a random effects model to provide a more
43 conservative estimate of the effect.

44 When no plausible explanation for the heterogeneity could be found, the quality of
45 the evidence was downgraded in GRADE for inconsistency.

1 *Assessing indirectness in intervention reviews*

2 Directness refers to the extent to which populations, interventions, comparisons and
3 outcomes reported in the evidence are similar to those defined in the inclusion
4 criteria for the review and was assessed by comparing the PICO elements in the
5 studies to the PICO defined in the review protocol. Indirectness is important when
6 such differences are expected to contribute to a difference in effect size, or may
7 affect the balance of benefits and harms considered for an intervention.

8 *Assessing imprecision and clinical importance in intervention reviews*

9 Imprecision in GRADE methodology refers to uncertainty around the effect estimate
10 and whether or not there is a clinically important difference between interventions
11 (that is, whether the evidence clearly supports a particular recommendation or
12 appears to be consistent with several candidate recommendations). Therefore,
13 imprecision differs from other aspects of evidence quality because it is not concerned
14 with whether the point estimate is accurate or correct (has internal or external
15 validity). Instead, it is concerned with uncertainty about what the point estimate
16 actually represents. This uncertainty is reflected in the width of the CI.

17 The 95% CI is defined as the range of values within which the population value will
18 fall on 95% of repeated samples, were the procedure to be repeated. The larger the
19 study, the smaller the 95% CI will be and the more certain the effect estimate.

20 Imprecision was assessed in the guideline evidence reviews by considering whether
21 the width of the 95% CI of the effect estimate was relevant to decision making,
22 considering each outcome independently. This is illustrated in Figure 1, which
23 considers a positive outcome for the comparison of treatment 'A' versus treatment
24 'B'. Three decision-making zones can be differentiated, bounded by the thresholds
25 for clinical importance (minimally important differences; MIDs) for benefit and harm.
26 The MID for harm for a positive outcome means the threshold at which treatment A is
27 less effective than treatment B by an amount that is clinically important to people with
28 the condition of interest (favours B).

29 When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no
30 uncertainty about the size and direction of effect, therefore, the effect estimate is
31 considered precise; that is, there is no imprecision.

32 When the CI crosses 2 zones, it is uncertain in which zone the true value of the effect
33 estimate lies and therefore there is uncertainty over which decision to make. The CI
34 is consistent with 2 possible decisions, therefore, the effect estimate is considered to
35 be imprecise in the GRADE analysis and the evidence is downgraded by 1 level
36 ('serious imprecision').

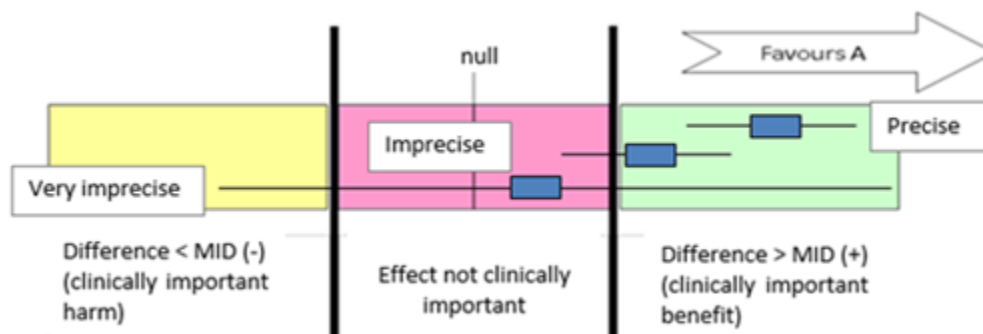
37 When the CI crosses all 3 zones, the effect estimate is considered to be very
38 imprecise because the CI is consistent with 3 possible clinical decisions and there is
39 therefore a considerable lack of confidence in the results. The evidence is therefore
40 downgraded by 2 levels in the GRADE analysis ('very serious imprecision').

41 Implicitly, assessing whether a CI is in, or partially in, a clinically important zone,
42 requires the guideline committee to estimate an MID or to say whether they would
43 make different decisions for the 2 confidence limits.

44 For most outcomes, however, there was a lack of an agreed MID. Instead outcomes
45 with low numbers of events or small sample size were downgraded for imprecision
46 using the following criteria

- 1 • For dichotomous outcomes: downgrade by 1 if <300 events altogether
2 (across both arms)
3 • For continuous outcomes: downgrade by 1 if sample size <400 altogether
4 (across both arms)

5 **Figure 1: Assessment of imprecision and clinical importance in intervention**
6 **reviews using GRADE**



7
8 MID: minimally important difference

9 *Defining minimally important differences for intervention reviews*

10 The committee was asked whether there were any recognised or acceptable MID in the clinical literature and community relevant to the review questions under
11 consideration. The following MID for quality of life measures were identified in the
12 clinical literature and approved by the committee:
13

- 14 • EORTC QLQ-C30: 5 points
15 • EORTC QLQ-CR29: 5 points
16 • EORTC QLQ-CR38: 5 points
17 • EQ-5D: 0.09 using FACT-G quintiles
18 • FACT-C: 5 points
19 • FACT-G: 5 points*
20 • SF-12: > 3.77 for the mental component summary (MCS) and > 3.29 for the
21 physical component summary (PCS) of the Short Form SF-12 (SF-12)
22 • SF-36: > 7.1 for the physical functioning scale, > 4.9 for the bodily pain scale,
23 and > 7.2 for the physical component summary.

24 For other outcomes (for example, overall survival, disease-free survival or
25 complication), any statistically significant change was considered by the committee to
26 of potential clinical importance.

27 **Prognostic studies**

28 ***Adapted GRADE methodology for prognostic reviews***

29 For prognostic reviews with evidence from comparative studies an adapted GRADE
30 approach was used. As noted above, GRADE methodology is designed for
31 intervention reviews but the quality assessment elements were adapted for
32 prognostic reviews.

33 The evidence for each outcome in the prognostic reviews was examined separately
34 for the quality elements listed and defined in Table 5. The criteria considered in the
35 rating of these elements are discussed below. Each element was graded using the

1 quality levels summarised in Table 3. Footnotes to GRADE tables were used to
 2 record reasons for grading a particular quality element as having ‘serious’ or ‘very
 3 serious’ quality issues. The ratings for each component were combined to obtain an
 4 overall assessment of quality for each outcome as described in Table 4.

5 **Table 5: Adaptation of GRADE quality elements for prognostic reviews**

Quality element	Description
Risk of bias (‘Study limitations’)	Limitations in study design and implementation may bias estimates and interpretation of the effect of the prognostic/risk factor. High risk of bias for the majority of the evidence reduces confidence in the estimated effect. Prognostic studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)
Inconsistency	This refers to unexplained heterogeneity between studies looking at the same prognostic/risk factor, resulting in wide variability in estimates of association (such as RRs or ORs), with little or no overlap in confidence intervals
Indirectness	This refers to any departure from inclusion criteria listed in the review protocol (such as differences in study populations or prognostic/risk factors), that may affect the generalisability of results
Imprecision	This occurs when a study has relatively few participants, a low event rate and also when the number of participants is too small for a multivariable analysis (as a rule of thumb, 10 participants are needed per variable).

6 *RR: relative risk; OR: odds ratio*

7 *Assessing risk of bias in prognostic reviews*

8 The Quality in Prognosis Studies (QUIPS) tool developed by Hayden 2013 was used
 9 to assess risk of bias in studies included in prognostic reviews (see Appendix H in
 10 the [Developing NICE guidelines: the manual](#); NICE 2014). The risk of bias in each
 11 study was determined by assessing the following domains:

- 12 • selection bias
- 13 • attrition bias
- 14 • prognostic factor bias
- 15 • outcome measurement bias
- 16 • control for confounders
- 17 • appropriate statistical analysis.

18 *Assessing inconsistency in prognostic reviews*

19 Where multiple results were deemed appropriate to meta-analyse (i.e. sufficient
 20 similarity between risk factor and outcome under investigation) inconsistency was
 21 assessed by visually inspecting forest plots and observing whether there was
 22 considerable heterogeneity in the results of the meta-analysis. This was assessed by
 23 calculating the I-squared statistic for the meta-analysis with an I-squared value of
 24 more than 50% indicating considerable heterogeneity, and more than 80% indicating
 25 very serious heterogeneity. When considerable or very serious heterogeneity was
 26 observed, possible reasons were explored and subgroup analyses were performed
 27 as pre-specified in the review protocol where possible.

1 When no plausible explanation for the heterogeneity could be found, the quality of
2 the evidence was downgraded in GRADE for inconsistency.

3 If meta-analysis was not appropriate (for example due to clinical heterogeneity)
4 inconsistency was assessed by visually inspecting forest plots.

5 *Assessing indirectness in prognostic reviews*

6 Indirectness in prognostic reviews was assessed by comparing the populations,
7 prognostic factors and outcomes in the evidence to those defined in the review
8 protocol.

9 *Assessing imprecision and clinical importance in prognostic reviews*

10 The committee considered any statistically significant change to be of potential
11 clinical importance. Outcomes were downgraded for imprecision if the overall sample
12 size was considered too small for multivariable analysis (using a rule of thumb of 10
13 participants needed per included variable) or if there were fewer than 300 events for
14 dichotomous outcomes.

15 **Qualitative reviews**

16 ***Adapted GRADE-CERQual methodology for qualitative reviews***

17 For the qualitative review an adapted GRADE Confidence in the Evidence from
18 Reviews of Qualitative research (GRADE-CERQual) approach (Lewin 2015) was
19 used. In this approach the quality of evidence is considered according to themes in
20 the evidence. The themes may have been identified in the primary studies or they
21 may have been identified by considering the reports of a number of studies. Quality
22 elements assessed using GRADE-CERQual are listed and defined in Table 6. Each
23 element was graded using the levels of concern summarised in Table 7. The ratings
24 for each component were combined (as with other types of evidence) to obtain an
25 overall assessment of quality for each theme as described in Table 8.

26 **Table 6: Adaptation of GRADE quality elements for qualitative reviews**

Quality element	Description
Risk of bias ('Methodological limitations')	Limitations in study design and implementation may bias interpretation of qualitative themes identified. High risk of bias for the majority of the evidence reduces confidence in review findings. Qualitative studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)
Relevance (or applicability) of evidence	This refers to the extent to which the evidence supporting the review findings is applicable to the context specified in the review question
Coherence of findings	This refers to the extent to which review findings are well grounded in data from the contributing primary studies and provide a credible explanation for patterns identified in the evidence
Adequacy of data (theme saturation or sufficiency)	This corresponds to a similar concept in primary qualitative research, that is, whether a theoretical point of theme saturation was achieved, at which point no further citations or observations would provide more insight or suggest a different interpretation of the particular theme. Individual studies that may have contributed to a theme or sub-theme may have been conducted in a manner that by design would have not reached theoretical saturation at an individual study level

1 **Table 7: CERQual levels of concern (by quality element)**

Level of concern	Definition
None or very minor concerns	Unlikely to reduce confidence in the review finding
Minor concerns	May reduce confidence in the review finding
Moderate concerns	Will probably reduce confidence in the review finding
Serious concerns	Very likely to reduce confidence in the review finding

2 **Table 8: Overall confidence in the evidence in CERQual (by review finding)**

Overall confidence level	Definition
High	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest
Moderate	It is likely that the review finding is a reasonable representation of the phenomenon of interest
Low	It is possible that the review finding is a reasonable representation of the phenomenon of interest
Very low	It is unclear whether the review finding is a reasonable representation of the phenomenon of interest

3 *Assessing methodological limitations in qualitative reviews*

4 Methodological limitations in qualitative studies were assessed using the Critical
5 Appraisal Skills Programme (CASP) checklist for qualitative studies (see appendix H
6 in [Developing NICE guidelines: the manual](#); NICE 2014). Overall methodological
7 limitations was derived by assessing the 6 domains summarised in Table 9.

8 **Table 9: Methodological limitations in qualitative studies**

Aim and appropriateness of qualitative evidence	This domain assesses whether the aims and relevance of the study were described clearly and whether qualitative research methods were appropriate for investigating the research question
Rigour in study design or validity of theoretical approach	This domain assesses whether the study approach was documented clearly and whether it was based on a theoretical framework (such as ethnography or grounded theory). This does not necessarily mean that the framework has to be stated explicitly, but a detailed description ensuring transparency and reproducibility should be provided
Sample selection	This domain assesses the background, the procedure and reasons for the method of selecting participants. The assessment should include consideration of any relationship between the researcher and the participants, and how this might have influenced the findings
Data collection	This domain assesses the documentation of

	the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses who conducted any interviews, how long they lasted and where they took place
Data analysis	This domain assesses whether sufficient detail was documented for the analytical process and whether it was in accordance with the theoretical approach. For example, if a thematic analysis was used, the assessment would focus on the description of the approach used to generate themes. Consideration of data saturation would also form part of this assessment (it could be reported directly or it might be inferred from the citations documented that more themes could be found)
Results	This domain assesses any reasoning accompanying reporting of results (for example, whether a theoretical proposal or framework is provided)

1 *Assessing relevance of evidence in qualitative reviews*

2 Relevance (applicability) of findings in qualitative research is the equivalent of
3 indirectness for quantitative outcomes, and refers to how closely the aims and
4 context of studies contributing to a theme reflect the objectives outlined in the
5 guideline review protocol.

6 *Assessing coherence of findings in qualitative reviews*

7 For qualitative research, a similar concept to inconsistency is coherence, which
8 refers to the way findings within themes are described and whether they make sense.
9 This concept was used in the quality assessment across studies for individual
10 themes. This does not mean that contradictory evidence was automatically
11 downgraded, but that it was highlighted and presented, and that reasoning was
12 provided. Provided the themes, or components of themes, from individual studies fit
13 into a theoretical framework, they do not necessarily have to reflect the same
14 perspective. It should, however, be possible to explain these by differences in context
15 (for example, the views of healthcare professionals might not be the same as those
16 of family members, but they could contribute to the same overarching themes).

17 *Assessing adequacy of data in qualitative reviews*

18 Adequacy of data (theme saturation or sufficiency) corresponds to a similar concept
19 in primary qualitative research in which consideration is made of whether a
20 theoretical point of theme saturation was achieved, meaning that no further citations
21 or observations would provide more insight or suggest a different interpretation of the
22 theme concerned. As noted above, it is not equivalent to the number of studies
23 contributing to a theme, but rather to the depth of evidence and whether sufficient
24 quotations or observations were provided to underpin the findings.

25 *Assessing clinical importance in qualitative reviews*

26 For themes stemming from qualitative findings, clinical importance was agreed by the
27 committee taking account of the generalisability of the context from which the theme

1 was derived and whether it was sufficiently convincing to support or warrant a
2 change in current practice, as well as the quality of the evidence.

3 **Reviewing economic evidence**

4 **Inclusion and exclusion of economic studies**

5 The titles and abstracts of papers identified through the searches were independently
6 assessed for inclusion using predefined eligibility criteria defined in Table 10.

7 **Table 10: Inclusion and exclusion criteria for the systematic reviews of** 8 **economic evaluations**

Inclusion criteria
Economic evaluations that compare costs and health consequences of interventions (i.e. true cost-effectiveness analyses)
Population, interventions, comparators and outcomes match those specified in the PICO
Quality of life based outcomes were used as the measure of effectiveness in at least one of the analyses presented
Incremental results reported or enough information for incremental results to be derived
Conducted from the perspective of a healthcare system in an OECD country
Exclusion criteria
Conference abstracts with insufficient methodological details for quality assessment
Non-English language papers

9 *OECD: Organisation for Economic Co-operation and Development; PICO: Population, Intervention,*
10 *Comparison, and Outcome*

11 Once the screening of titles and abstracts was complete, full versions of the selected
12 papers were acquired for assessment. The quality of evidence was assessed using
13 the economic evaluations checklist as specified in [Developing NICE guidelines: the](#)
14 [manual](#) (NICE 2014).

15 **Appraising the quality of economic evidence**

16 The quality of economic evaluations in this guideline were appraised using the
17 methodology checklist reported in the [Developing NICE guidelines: the manual](#),
18 Appendix H (NICE 2014) for all studies which met the inclusion criteria.

19 **Health economic modelling**

20 The aims of the health economic input to the guideline were to inform the guideline
21 committee of potential economic issues related to primary colorectal cancer and
22 colorectal cancer metastases in adults to ensure that recommendations represented
23 a cost-effective use of healthcare resources. Health economic evaluations aim to
24 integrate data on healthcare benefits (ideally in terms of quality-adjusted life-years,
25 QALYs) with the costs of different care options. In addition, the health economic input
26 aimed to identify areas of high resource impact; recommendations which, while
27 nevertheless cost-effective, might have a large impact on Clinical Commissioning
28 Group or Trust finances and so need special attention.

29 The committee prioritised 2 economic models on: the treatment of liver metastases
30 which were amenable to treatment with curative intent and surgical techniques for

1 rectal cancer. The committee thought economic considerations would be particularly
2 important when formulating recommendations in these areas.

3 The methods and results of the de novo economic analyses are reported in Appendix
4 J of the relevant evidence reports. When new economic analysis was not prioritised,
5 the committee made a qualitative judgement regarding cost effectiveness by
6 considering expected differences in resource and cost use between options,
7 alongside clinical effectiveness evidence identified from the clinical evidence review.

8 **Cost effectiveness criteria**

9 NICE's report [Social value judgements: principles for the development of NICE](#)
10 [guidance](#) sets out the principles that committees should consider when judging
11 whether an intervention offers good value for money. In general, an intervention was
12 considered to be cost effective if any of the following criteria applied (given that the
13 estimate was considered plausible):

- 14 • the intervention dominated other relevant strategies (that is, it was both less
15 costly in terms of resource use and more clinically effective compared with all
16 the other relevant alternative strategies), or
- 17 • the intervention cost less than £20,000 per QALY gained compared with the
18 next best strategy, or
- 19 • the intervention provided clinically important benefits at an acceptable
20 additional cost when compared with the next best strategy.

21 The committee's considerations of cost-effectiveness are discussed explicitly under
22 the 'Cost Effectiveness and Resource Use' headings of the relevant sections.

23 **Expert evidence**

24 When the guideline developers believe that there is important evidence in addition to
25 that identified by the literature searches, they can call upon relevant individuals or
26 organisations to provide expert evidence to the guideline committee.

27 When agreeing the review protocols, the committee considered areas that might
28 benefit from expert evidence. The committee were aware of 2 randomised trials due
29 to publish just before the guideline itself but not soon enough to be included in the
30 evidence review. They agreed that the results of these trials could change current
31 practice and invited the trialists to provide expert evidence. In both cases this
32 evidence was provided orally and was discussed and considered by the committee.
33 A written summary of the expert evidence was also included as an appendix to both
34 evidence reports.

35 The topics covered by expert evidence were:

- 36 • neoadjuvant systemic anticancer therapy for operable colon cancer
- 37 • stenting for large bowel obstruction caused by colorectal cancer.

38

39 A call for evidence to all registered stakeholders was not made because the
40 committee did not believe there was relevant unpublished evidence other areas,
41 beyond the two topics identified above. Unpublished evidence in the form of
42 conference abstracts was not included because such abstracts typically do not have
43 enough detail to be critical appraised.

1 **Developing recommendations**

2 **Guideline recommendations**

3 Recommendations were drafted on the basis of the committee's interpretation of the
4 available evidence, taking account of the balance of benefits, harms and costs
5 between different courses of action. When clinical and economic evidence was of
6 poor quality, conflicting or absent, the committee drafted recommendations based on
7 their expert opinion. The considerations for making consensus-based
8 recommendations include the balance between potential benefits and harms, the
9 economic costs or implications compared with the economic benefits, current
10 practices, recommendations made in other relevant guidelines, preferences of people
11 and equality issues.

12 The main considerations specific to each recommendation are outlined under the
13 heading 'The committee's discussion of the evidence' within each evidence report.

14 For further details refer to [Developing NICE guidelines: the manual](#) (NICE 2014).

15 **Research recommendations**

16 When areas were identified for which evidence was lacking, the committee
17 considered making recommendations for future research. For further details refer to
18 [Developing NICE guidelines: the manual](#) (NICE 2014).

19 **Validation process**

20 This guideline was subject to a 6-week public consultation and feedback process. All
21 comments received from registered stakeholders were responded to in writing and
22 posted on the NICE website at publication. For further details refer to [Developing](#)
23 [NICE guidelines: the manual](#) (NICE 2014).

24 **Updating the guideline**

25 Following publication, NICE will undertake a surveillance review to determine
26 whether the evidence base has progressed sufficiently to consider altering the
27 guideline recommendations and warrant an update. For further details refer to
28 [Developing NICE guidelines: the manual](#) (NICE 2014).

29 **Funding**

30 The NGA was commissioned by NICE to develop this guideline.

References

- 1
2 **Bradburn 2007**
3
4 Bradburn M, Deeks J, Berlin J, et al. (2007) Much ado about nothing: A comparison
5 of the performance of meta-analytical methods with rare events. *Statistics in*
6 *Medicine* 26(1): 53–77
- 7 **Hayden 2013**
8 Hayden J, van der Windt D, Cartwright L, et al. (2013) Assessing Bias in Studies of
9 Prognostic Factors. *Annals of Internal Medicine* 158(4): 280–286
- 10 **Higgins 2011**
11 Higgins J, Green S (editors) (2011) *Cochrane Handbook for Systematic Reviews of*
12 *Interventions* Version 5.1.0 [updated March 2011] The Cochrane Collaboration.
13 Available from www.handbook.cochrane.org (accessed 15 June 2018)
- 14 **Lewin 2015**
15 Lewin S, Glenton C, Munthe-Kaas, H et al. (2015) Using qualitative evidence in
16 decision making for health and social interventions: an approach to assess
17 confidence in findings from qualitative evidence syntheses (GRADE-CERQual).
18 *PLoS Med* 12(10): e1001895
- 19 **NICE 2014**
20 National Institute for Health and Care Excellence (NICE) (2019) [Developing NICE](#)
21 [guidelines: the manual](#) (updated 2018, accessed 15 June 2018)
- 22 **NICE 2018**
23 National Institute for Health and Care Excellence (NICE) (2014) [NICE Policy on](#)
24 [conflict of interest](#) (updated 2017, accessed 15 June 2018)
- 25 **Santesso 2016**
26 Santesso N, Carrasco-Labra A, Langendam M, et al. (2016) Improving GRADE
27 evidence tables part 3: detailed guidance for explanatory footnotes supports creating
28 and understanding GRADE certainty in the evidence judgments. *Journal of Clinical*
29 *Epidemiology* 74: 28-39
- 30 **Tierney 2007**
31 Tierney J, Stewart L, Ghersi D, et al. (2007) Practical methods for incorporating
32 summary time-to-event data into meta-analysis. *Trials* 8: 1-16
- 33 **Viechtbauer 2010**
34 Viechtbauer W (2010). Conducting meta-analyses in R with the metafor package.
35 *Journal of Statistical Software* 36(3): 1-48