National Institute for Health and Care Excellence

Draft for consultation

Acne vulgaris: management

Methods

NICE guideline tbc

Supplement 1

December 2020

Draft for consultation

Supplementary material was developed by the National Guideline Alliance which is a part of 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN:

Contents

Development of the guideline

Remit

The National Institute for Health and Care Excellence (NICE) commissioned the National Guideline Alliance (NGA) to develop a guideline for the management of acne vulgaris.

For further details of what the guideline does and does not cover see: https://www.nice.org.uk/guidance/gid-ng10109/documents/final-scope.

Methods

Introduction

This section summarises methods used to identify and review the evidence, to consider cost effectiveness, and to develop guideline recommendations. This guideline was developed in accordance with methods described in Developing NICE guidelines: the manual (NICE 2018a).

Declarations of interest were recorded and managed in accordance with NICE's 2018 Policy on declaring and managing interests for NICE advisory committees (NICE 2018b).

Developing the review questions and outcomes

The review questions considered in this guideline were based on the key areas identified in the guideline scope .They were drafted by the NGA technical team, and refined and validated by the guideline committee.

The review questions were based on the following frameworks:

- intervention reviews – using population, intervention, comparison and outcome (PICO)
- prognostic reviews – using population, presence or absence of a prognostic, risk or predictive factor and outcome (PPO)
- qualitative reviews – using population, phenomenon of interest and context

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

Literature searches, critical appraisal and evidence reviews were completed for all review questions.

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

Table 1: Summary of review questions and index to evidence reviews

Question number	Evidence review ID*	Review question	Type of review
Individual topics: Evidence reviews covering one review question			
1.	A	What information and support is valued by people with acne vulgaris, and their parents or carers?	Qualitative
2.	B	What skin cleansing advice is effective in the treatment of acne vulgaris?	Intervention
3.	C	What is the effectiveness of dietary interventions for acne vulgaris, for example • milk free diet • dairy product free diet	Intervention

		• low glycaemic load diet?	
4.	D	When should people with acne vulgaris be referred to specialist care?	Intervention
5.	J	Is the addition of oral corticosteroids to oral isotretinoin of benefit for the treatment of severe acne (including acne conglobata and acne fulminans)?	Intervention
6.	K	What is the effectiveness of intralesional corticosteroids in the treatment of individual acne vulgaris lesions?	Intervention
7.	L	What are the risk factors for scarring in people with acne vulgaris?	Prognostic
8.	M	What interventions are effective in the management of scarring resulting from acne vulgaris, for example • microneedling techniques • laser treatment • intradermal injection (for example, autologous platelet-rich plasma; autologous fibroblasts; polymethylmethacrylate (PMMA) microspheres in collagen) • surgical treatment (for example, subcuticular incision)?	Intervention
Combined topics: Evidence reviews covering more than one question **Questions 9-17 below are covered by 5 overarching review questions:** **For people with mild to moderate acne vulgaris what are the most effective treatment options? (E1 refers to network meta-analysis and E2 refers to the pairwise meta-analysis of treatment options)** **For people with moderate to severe acne vulgaris what are the most effective treatment options? (F1 refers to network meta-analysis and F2 refers to the pairwise meta-analysis of treatment options)** **What is an effective management option for people with acne vulgaris and polycystic ovary syndrome (PCOS)? (G)** **What is the effectiveness of topical or oral pharmacological and physical interventions in treatment resistant acne vulgaris? (H)** **What is the effectiveness of topical or oral pharmacological and physical maintenance treatment for acne vulgaris? (I)**			
9.	E1/E2 F1/F2 G, H, I	What is the effectiveness of topical treatments individually or in combination in the treatment of acne vulgaris , for example: • benzoyl peroxide • antibiotics • antiseptics • retinoids and retinoid-like agents (for example, tretinoin, adapalene)	

		• azelaic acid • nicotinamide • combination of antibiotic and retinoid or retinoid-like agent • combination of benzoyl peroxide and retinoid or retinoid-like agent • combination of antibiotic and benzoyl peroxide?	Intervention
10.	E1/E2 F1/F2 G, H, I	What is the effectiveness of oral antibiotic treatments individually or in combination in the treatment of acne vulgaris, for example: • tetracyclines (for example oxytetracycline, doxycycline, minocycline, tetracycline, lymecycline) • macrolide antibiotics (for example, erythromycin and azithromycin) • trimethoprim?	
11.	E1/E2 F1/F2 G, H, I	What is the effectiveness of an oral antibiotic with a topical agent compared to oral antibiotic alone in the treatment of acne vulgaris?	
12.	E1/E2 F1/F2 G, H, I	What is the optimal duration of antibiotic treatments (topical and systemic) for acne vulgaris?	
13.	E1/E2 F1/F2 G, H, I	What is the effectiveness of hormonal contraceptive treatments for people with acne vulgaris?	
14.	E1/E2 F1/F2 G, H, I	What is the effectiveness of non- hormonal contraceptive anti-androgens (including spironolactone) in the treatment of acne vulgaris?	
15.	E1/E2 F1/F2 G, H, I	What is the effectiveness of metformin in people with acne vulgaris?	
16.	E1/E2 F1/F2 G, H, I	What is the effectiveness of oral isotretinoin for acne vulgaris?	
17.	E1/E2 F1/F2 G, H, I	What is the effectiveness of physical treatments for acne vulgaris, for example • comedone extraction • chemical peels (for example, glycolic acid, lactic acid, salicylic acid) • intralesional steroids • light devices (for example, intense pulsed light, photopneumatic therapy and photodynamic therapy)?	

** this refers to the alphabetical or alphanumeric ID of evidence reviews in the guideline*

The COMET database was searched for core outcome sets relevant to this guideline. No core outcome sets were identified and therefore the outcomes were chosen based on committee discussions.

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

- Supplement 2 – NGA team and collaborators from the TSU
- Supplement 3 – TSU NMA software code (mild to moderate acne)
- Supplement 4 – NMA data (mild to moderate acne)
- Supplement 5 – NMA of efficacy: included and excluded studies (mild to moderate acne)
- Supplement 6 – NMA, direct and indirect estimates (mild to moderate acne)
- Supplement 7 – TSU NMA software code (moderate to severe acne)
- Supplement 8 – NMA data (moderate to severe acne)
- Supplement 9 – NMA of efficacy: included and excluded studies (moderate to severe acne)
- Supplement 10 – NMA, direct and indirect estimates (moderate to severe acne)

Searching for evidence

Scoping search

During the scoping phase, searches were conducted for previous guidelines, economic evaluations, health technology assessments and randomized controlled trials and systematic reviews. Searches of websites of organisations and institutional repositories were also undertaken for relevant documents. Any references suggested by stakeholders at the scoping consultation were considered.

Systematic literature search

Systematic literature searches were undertaken to identify published evidence relevant to each review question.

Databases were searched using subject headings, free-text terms and, where appropriate, study type filters. Where possible, searches were limited to retrieve studies published in English. All searches were conducted in Medline, Embase, Cochrane Central Register of Controlled Trials (CCTR) and Cochrane Database of Systematic Reviews (CDSR). For the review question related to Q13.1, CINAHL was also searched.

Searches were run once for all reviews during development. Searches for the following questions were updated in May 2020.

- Effectiveness of topical or oral pharmacological and physical interventions in the treatment of acne vulgaris

Details of the search strategies, including the study-design filters used and databases searched, are provided in Appendix B of each evidence review.

Economic systematic literature search

Systematic literature searches were also undertaken to identify published economic evidence. An additional search was undertaken of ‘studies reporting health state utility data that could be utilised in a cost-utility analysis’. Databases were searched using subject headings, free-text terms and, where appropriate, an economic evaluations search filter.

A single search, using the population search terms used in the evidence reviews, was conducted to identify economic evidence in the NHS Economic Evaluation Database (NHS EED) and Health Technology Assessments (HTA). Another single search, using the population search terms used in the evidence reviews combined with an economic evaluations search filter and additionally, a health state utility search filter, was conducted in Medline, Embase, the Cochrane Central Register of Controlled Trials (CCTR). Where possible, searches were limited to studies published in English.

As with the general literature searches, the economic literature searches were updated in May 2020.

Details of the search strategies, including the study-design filter used and databases searched, are provided in Appendix B of each evidence review.

Quality assurance

Search strategies were quality assured by cross-checking reference lists of relevant studies, analysing search strategies from published systematic reviews and asking members of the committee to highlight key studies. The principal search strategies for each search were also quality assured by a second information scientist using an adaptation of the PRESS 2015 Guideline Evidence-Based Checklist (McGowan 2016). In addition, all publications highlighted by stakeholders at the time of the consultation on the draft scope were considered for inclusion.

Reviewing evidence

Systematic review process

The evidence was reviewed in accordance with the following approach.

- Potentially relevant articles were identified from the search results for each review question by screening titles and abstracts. Full-text copies of the articles were then obtained.
- Full-text articles were reviewed against pre-specified inclusion and exclusion criteria in the review protocol (see Appendix A of each evidence review).
- Key information was extracted from each article on study methods and results, in accordance with factors specified in the review protocol. The information was presented in a summary table in the corresponding evidence review and in a more detailed evidence table (see Appendix E of each evidence review).
- Included studies were critically appraised using an appropriate checklist as specified in Developing NICE guidelines: the manual (NICE 2018a). Further detail on appraisal of the evidence is provided below.
- Summaries of evidence by outcome were presented in the corresponding evidence review and discussed by the committee.

Review questions informing network meta-analyses (NMA) were subject to dual screening, study selection and data extraction. Other review questions selected as high priorities for economic analysis (and those selected as medium priorities and where economic analysis could influence recommendations), were subject to dual screening and study selection through a 10% random sample of articles. Any discrepancies were resolved by discussion between the first and second reviewers or

by reference to a third (senior) reviewer. For the remaining review questions, internal (NGA) quality assurance processes included consideration of the outcomes of screening, study selection and data extraction and the committee reviewed the results of study selection and data extraction.

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

Type of studies and inclusion/exclusion criteria

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

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

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

Topical and physical treatments for acne vulgaris are sometimes tested in split-face trials, where the right and left sides of the same person's face are randomly allocated different treatments. Due to unit of analysis issues, results from such trials were only included if they were presented as the difference in outcome between left and right sides of the face.

For prognostic reviews, prospective and retrospective cohort and case–control studies and case series were considered for inclusion. Studies that included multivariable analysis were prioritised.

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

The committee was consulted about any uncertainty regarding inclusion or exclusion of studies. A list of excluded studies for each review question, including reasons for exclusion is presented in Appendix D of the corresponding evidence review.

Narrative reviews, posters, letters, editorials, comment articles, unpublished studies and studies published in languages other than English were excluded. Conference abstracts were not considered for inclusion because conference abstracts typically do not have sufficient information to allow for full critical appraisal.

Methods of combining evidence

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

Data synthesis for intervention reviews

Pairwise meta-analysis

Meta-analysis to pool results from RCTs was conducted where possible using Cochrane Review Manager (RevMan5) software. Where non-randomised evidence was used, this was/was not meta-analysed.

For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a fixed effect model was used to calculate risk ratios (RRs). For all outcomes with zero events in both arms the risk difference was presented. For outcomes in which the majority of studies had low event rates (<1%), Peto odds ratios (ORs) were calculated as this method performs well when events are rare (Bradburn 2007).

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

If a study reported only the summary statistic and 95% CI the generic-inverse variance method was used to enter data into RevMan5. If the control event rate was reported this was used to generate the absolute risk difference in GRADEpro. If multivariable analysis was used to derive the summary statistic but no adjusted control event rate was reported, no absolute risk difference was calculated.

When evidence was based on studies that reported descriptive data or medians with interquartile ranges or p values, this information was included in the corresponding GRADE tables (see below) without calculating relative or absolute effects. Consequently, certain aspects of quality assessment such as imprecision of the effect estimate could not be assessed as per standard methods for this type of evidence and subjective ratings were considered instead.

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

When meta-analysis was undertaken, the results were presented visually using forest plots generated using RevMan5 (see Appendix F of relevant evidence reviews).

When case series were included, descriptive data from the studies were included and no further analysis was performed.

Network meta-analysis

Network meta-analysis (NMA) is a generalisation of standard pairwise meta-analysis for A versus B trials, to data structures that include, for example, A versus B, B versus C, and A versus C trials (Dias 2011; Lu 2004). A basic assumption of NMA methods is that direct and indirect evidence estimate the same parameter, that is, the relative effect between A and B measured directly from an A versus B trial, is the same with the relative effect between A and B estimated indirectly from A versus C and B versus C trials. NMA techniques strengthen inference concerning the relative effect of two treatments by including both direct and indirect comparisons between treatments, and, at the same time, allow simultaneous inference on all treatments examined in the pair-wise trial comparisons, which is essential for consideration of

treatment in economic analysis (Caldwell 2005; Lu 2004). Simultaneous inference on the relative effect of a number of treatments is possible provided that treatments participate in a single “network of evidence”, that is, every treatment is linked to at least one of the other treatments under assessment through direct or indirect comparisons. NMA takes all trial information into consideration, without ignoring part of the evidence and without introducing bias by breaking the rules of randomisation.

As is the case for ordinary pairwise meta-analysis, NMA may be conducted using either fixed or random effect models. A fixed effect model typically assumes that there is no variation in relative effects across trials for a particular pairwise comparison and any observed differences are solely due to chance. For a random effects model, it is assumed that the relative effects are different in each trial but that they are from a single common distribution. The variance reflecting heterogeneity is often assumed to be constant across trials.

Class models were used so that strength could be borrowed across treatments in the same class and to reconnect disconnected networks. Classes of treatments are groups of interventions which are thought to have similar modes of action and, consequently, similar effects. For all outcomes, both fixed and random class effects models were fitted. The random class effects model assumes the relative effects of treatments within a class are exchangeable. Treatment effects are shrunk towards a class mean and can borrow strength from other elements of the class. The fixed class effects model assumes treatments within a class have identical relative effects.

In a Bayesian analysis, for each parameter the evidence distribution is weighted by a distribution of prior beliefs. The Markov chain Monte Carlo (MCMC) algorithm was used to generate a sequence of samples from a joint posterior distribution of 2 or more random variables and is particularly well adapted to sampling the treatment effects (known as a posterior distribution) of a Bayesian network. A prior distribution was used to maximise the weighting given to the data and to generate the posterior distribution of the results.

For the analyses, a series of burn-in simulations were run to allow the posterior distributions to converge and then further simulations were run to produce the posterior outputs. Convergence was assessed by examining the history, autocorrelation and Brooks-Gelman-Rubin plots.

Goodness-of-fit of the models were also estimated by using the posterior mean of the sum of the deviance contributions for each item by calculating the residual deviance and the deviance information criterion (DIC). If the residual deviance was close to the number of unconstrained data points (the number of trial arms in the analysis) then the model was explaining the data at a satisfactory level. The choice of a fixed effect or random effects model can be made by comparing their goodness-of-fit to the data. Treatment specific posterior effects were generated for every possible pair of comparisons by combining direct and indirect evidence in each network.

Evidence of treatment effect was demonstrated if the 95% credible intervals [CrI] of the effect versus placebo did not cross the line of no effect.

For the outcome of efficacy, new models were developed by the NICE Guidelines Technical Support Unit, University of Bristol (TSU). For other outcomes, standard fixed and random effects models were adapted, available from NICE Decision Support Unit (DSU) technical support document number 2 (Dias 2011).

The NMA work was undertaken by the NICE Guidelines Technical Support Unit, University of Bristol (TSU).

Details of the NMA methods employed in this guideline are provided in appendix M of evidence reports E1 and F1.

Data synthesis for prognostic reviews

ORs or RRs with 95% CIs reported in published studies were extracted or calculated by the NGA technical team to examine relationships between risk factors and outcomes of interest. Ideally analyses would have adjusted for key confounders (such as age or parity) to be considered for inclusion. Recognising variation across studies in terms of populations, risk factors, outcomes and statistical analysis methods (including adjustments for confounding factors), prognostic data were not pooled, but results from individual studies were presented in the evidence reviews.

When case series were included, descriptive data from the studies were included and no further analysis was performed.

Data synthesis for qualitative reviews

Where possible, a meta-synthesis was conducted to combine evidence from qualitative studies. Whenever studies identified a qualitative theme relevant to the protocol, this was extracted and the main characteristics were summarised. When all themes had been extracted from studies, common concepts were categorised and tabulated. This included information on how many studies had contributed to each theme identified by the NGA technical team.

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

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

Appraising the quality of evidence

Intervention studies

Pairwise meta-analysis

GRADE methodology for intervention reviews

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

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

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

The initial quality rating was based on the study design: RCTs start as 'high' quality evidence, non-randomised studies start as 'low' quality evidence. The rating was then modified according to the assessment of each quality element (Table 2). Each quality element considered to have a 'serious' or 'very serious' quality issue was downgraded by 1 or 2 levels respectively (for example, evidence starting as 'high' quality was downgraded to 'moderate' or 'low' quality). In addition, there was a possibility to upgrade evidence from non-randomised studies (provided the evidence for that outcome had not previously been downgraded) if there was a large magnitude of effect, a dose–response gradient, or if all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results showed no effect.

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

Quality element	Description
Risk of bias ('Study limitations')	This refers to limitations in study design or implementation that reduce the internal validity of the evidence
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 few participants or few events of interest, resulting in wide confidence intervals that cross minimally 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

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

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

Assessing risk of bias in intervention reviews

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

Risk of bias in RCTs was assessed using the Cochrane risk of bias tool version 2 (see Appendix H in Developing NICE guidelines: the manual; NICE 2018a).

A study with a poor methodological design does not automatically imply high risk of bias; the bias is considered individually for each outcome and it is assessed whether the chosen design and methodology will impact on the estimation of the intervention effect.

More details about the Cochrane risk of bias tool version 2 can be found in Section 8 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2020).

For systematic reviews of RCTs the AMSTAR checklist was used and for systematic reviews of other study types the ROBIS checklist was used (see Appendix H in Developing NICE guidelines: the manual; NICE 2018a).

For non-randomised studies the ROBINS-I checklist was used (see Appendix H in Developing NICE guidelines: the manual; NICE 2018a).

Assessing inconsistency in intervention reviews

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

Inconsistency was assessed visually by inspecting forest plots and observing whether there was considerable heterogeneity in the results of the meta-analysis (for example if the point estimates of the individual studies consistently showed benefits or harms). This was supported by calculating the I-squared statistic for the meta-analysis with an I-squared value of more than 50% indicating serious heterogeneity, and more than 80% indicating very serious heterogeneity. When considerable or very serious heterogeneity was observed, possible reasons were explored and subgroup analyses were performed as pre-specified in the review protocol where possible. In the case of unexplained heterogeneity, sensitivity analyses were planned based on

the quality of studies, eliminating studies at high risk of bias (in relation to randomisation, allocation concealment and blinding, and/or missing outcome data).

When considerable heterogeneity was present, the meta-analysis was re-run using the Der-Simonian and Laird method with a random effects model and this was used for the final analysis.

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

Assessing indirectness in intervention reviews

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

Assessing imprecision and importance in intervention reviews

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

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

Imprecision was assessed in the guideline evidence reviews by considering whether the width of the 95% CI of the effect estimate was relevant to decision making, considering each outcome independently. This is illustrated in Figure 1, which considers a positive outcome for the comparison of two treatments. Three decision-making zones can be differentiated, bounded by the thresholds for minimal importance (minimally important differences; MIDs) for benefit and harm.

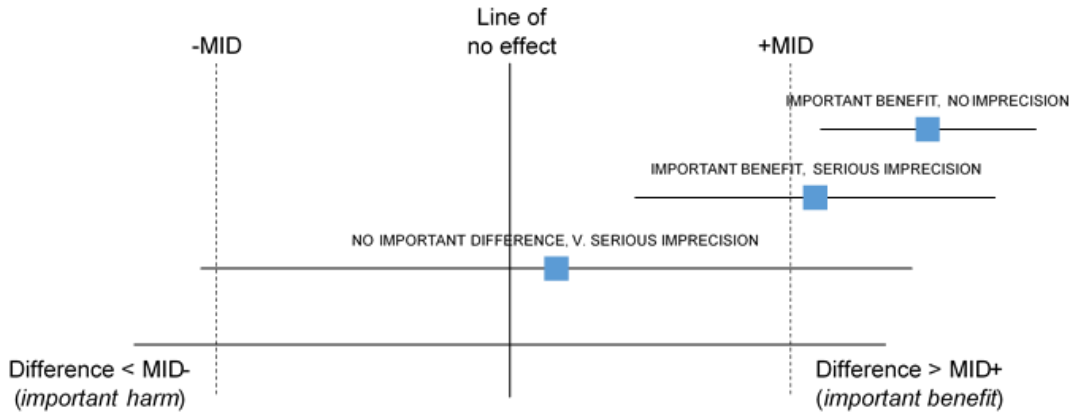
When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

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

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

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

Figure 1: Assessment of imprecision and importance in intervention reviews using GRADE



MID, minimally important difference

Defining minimally important differences for intervention reviews

The committee was asked whether there were any recognised or acceptable MIDs in the published literature and community relevant to the review questions under consideration. The committee was not aware of any MIDs that could be used for the guideline.

In the absence of published or accepted MIDs, the committee agreed to use the GRADE default MIDs to assess imprecision. For dichotomous outcomes minimally important thresholds for a RR of 0.8 and 1.25 respectively were used as default MIDs in the guideline. The committee also chose to use 0.8 and 1.25 as the MIDs for ORs & HRs in the absence of published or accepted MIDs. ORs were predominantly used in the guideline when Peto OR were indicated due to low event rates, at low event rates OR are mathematically similar to RR making the extrapolation appropriate. While no default MIDs exist for HR, the committee agreed for consistency to continue to use 0.8 and 1.25 for these outcomes.

If risk difference was used for meta-analysis, for example if the majority of studies had zero events in either arm, imprecision was assessed based on sample size using 300 and 500 as cut-offs for very serious and serious imprecision respectively. The committee used these numbers based on commonly used optimal information size thresholds.

The same thresholds were used as default MIDs in the guideline for all dichotomous outcomes considered in intervention evidence reviews. For continuous outcomes default MIDs are equal to half the median SD of the control groups at baseline (or at follow-up if the SD is not available a baseline).

Assessing publication bias in intervention reviews

Where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias. Where

fewer than 10 studies were included for an outcome, the committee subjectively assessed the likelihood of publication bias based on factors such as the proportion of trials funded by industry and the propensity for publication bias in the topic area.

Network meta-analysis

For the NMAs, quality was assessed by looking at risk of bias across the included evidence using the Cochrane Risk of Bias Tool for Randomized Controlled Trials, as well as heterogeneity and consistency (also called coherence). Heterogeneity concerns the differences in treatment effects between trials within each treatment contrast (measured by the posterior median between-study standard deviation and compared with treatment posterior mean effects), while consistency concerns the differences between the direct and indirect evidence informing the treatment contrasts. Inconsistency arises when there is a conflict between direct evidence (from an A vs. B trial) and indirect evidence (gained from A vs. C and B vs. C trials) and can only be assessed when there are closed loops of evidence on three treatments that are informed by at least three distinct trials (van Valkenhoef 2016).

To determine if there was evidence of inconsistency, in each analysis, the selected consistency model (fixed or random effects) was compared to an “inconsistency”, or unrelated mean effects, model. Further checks for evidence of inconsistency were performed through node-splitting (van Valkenhoef 2016).

Bias adjustment models were fitted to down-weight trials at high or unclear risk of bias for domains of the Cochrane Risk of Bias tool that had sufficient variability in the ratings. Models that adjusted for small study bias were also fitted (Dias 2010, Welton 2009).

Threshold analysis was undertaken to test the robustness of treatment recommendations based on the NMA, to potential biases or sampling variation in the included evidence. Threshold analysis has been developed as an alternative to GRADE for assessing confidence in guideline recommendations based on network meta-analysis (Phillippo 2018).

Prognostic studies

Adapted GRADE methodology for prognostic reviews

For prognostic reviews with evidence from comparative studies an adapted GRADE approach was used. As noted above, GRADE methodology is designed for intervention reviews but the quality assessment elements were adapted for prognostic reviews. Adapted GRADE was not used for evidence from case series; instead quality of case series evidence was assessed using the Checklist for Case Series developed by the Joanna Briggs Institute. More information about this tool can be found on the developer’s website.

The evidence for each outcome in the prognostic reviews was examined separately for the quality elements listed and defined in Table 5. The criteria considered in the rating of these elements are discussed below. Each element was graded using the quality levels summarised in Table 3. Footnotes to GRADE tables were used to record reasons for grading a particular quality element as having ‘serious’ or ‘very serious’ quality issues. The ratings for each component were combined to obtain an overall assessment of quality for each outcome as described in Table 4.

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 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). This was assessed by considering the confidence interval in relation to the point estimate for each outcome reported in the included studies

RR, relative risk; OR, odds ratio

Assessing risk of bias in prognostic reviews

The Quality in Prognosis Studies (QUIPS) tool developed by Hayden 2013 was used to assess risk of bias in studies included in prognostic reviews (see Appendix H in the Developing NICE guidelines: the manual; NICE 2018a). The risk of bias in each study was determined by assessing the following domains:

- selection bias
- attrition bias
- prognostic factor bias
- outcome measurement bias
- control for confounders
- appropriate statistical analysis.

Assessing inconsistency in prognostic reviews

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

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

Assessing indirectness in prognostic reviews

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

Assessing imprecision and importance in prognostic reviews

Prognostic studies may have a variety of purposes, for example, establishing typical prognosis in a broad population, establishing the effect of patient characteristics on prognosis, and developing a prognostic model. While by convention MIDs relate to intervention effects, the committee agreed to use GRADE default MIDs for intervention studies as a starting point from which to assess whether the size of an outcome effect in a prognostic study would be large enough to be meaningful in practice.

Qualitative reviews

GRADE-CERQual methodology for qualitative reviews

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

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

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

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

Assessing methodological limitations in qualitative reviews

Methodological limitations in qualitative studies were assessed using the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies (see appendix H in Developing NICE guidelines: the manual; NICE 2018a). Overall methodological limitations were derived by assessing the methodological limitations across the 6 domains summarised in Table 9.

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)

Assessing relevance of evidence in qualitative reviews

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

Assessing coherence of findings in qualitative reviews

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

Assessing adequacy of data in qualitative reviews

Adequacy of data (theme saturation or sufficiency) corresponds to a similar concept in primary qualitative research in which consideration is made of whether a theoretical point of theme saturation was achieved, meaning that no further citations or observations would provide more insight or suggest a different interpretation of the theme concerned. As noted above, it is not equivalent to the number of studies

contributing to a theme, but rather to the depth of evidence and whether sufficient quotations or observations were provided to underpin the findings.

Assessing importance in qualitative reviews

For themes stemming from qualitative findings, importance was agreed by the committee taking account of the generalisability of the context from which the theme was derived and whether it was sufficiently convincing to support or warrant a change in current practice, as well as the quality of the evidence.

Reviewing economic evidence

Systematic reviews of economic literature were conducted for all review questions covered in the guideline. In addition, literature on the health-related quality of life of the population covered by this guideline was systematically searched to identify studies reporting appropriate health state utility data that could be utilised in a cost-utility analysis.

Inclusion and exclusion of economic studies

Titles and abstracts of articles identified through the economic literature searches were assessed for inclusion using the predefined eligibility criteria listed in Table 10.

Table 10: Inclusion and exclusion criteria for systematic reviews of economic evaluations

Inclusion criteria
Only studies from Organisation for Economic Co-operation and Development member countries were included, as the aim of the review was to identify economic information transferable to the UK context.
Only studies published from 2000 onwards were included in the review. This date restriction was imposed so that retrieved economic evidence was relevant to current healthcare settings and costs.
Intervention or comparators in accordance with the guideline scope
Study population in accordance with the guideline scope
Full economic evaluations (cost-utility, cost effectiveness, cost-benefit or cost-consequence analyses) assessing both costs and outcomes associated with interventions of interest, as well as costing analyses that compared only costs between 2 or more interventions of interest were included in the review
Studies were included provided that sufficient details regarding methods and results were available to enable the methodological quality of the study to be assessed, and provided that the study's data and results were extractable.
Clinical effectiveness data utilised in the economic study should have been derived from a clinical trial, a prospective or retrospective cohort study, or from a literature review.
The outcome measure of the economic analysis should be the Quality Adjusted Life Year (QALY) or one of the measures considered in the clinical review.
Studies should be reporting separately costs from a healthcare (and, if available, personal social services) perspective.
Exclusion criteria
Poster presentations, conference abstracts and letters containing insufficient methodological details
Non-English language papers

Inclusion criteria
Cost-of-illness type studies
Non-comparative studies
Before-and-after studies and studies based on retrospective analyses of administrative healthcare data, due to associated methodological limitations and overall low quality characterising these study designs.
Studies that considered exclusively intervention costs, e.g. drug acquisition costs, without considering wider healthcare costs associated with the management of acne
Studies that compared costs of branded vs generic forms of the same drug

Once the screening of titles and abstracts was completed, full-text copies of potentially relevant articles were requested for detailed assessment. Inclusion and exclusion criteria were applied to articles obtained as full-text copies.

No economic studies met inclusion criteria for the review. Lists of economic studies excluded after obtaining full text with reasons for exclusion are provided in the appendix K of the relevant evidence reviews. The PRISMA for the search of economic evaluations is presented in the appendix G of each evidence review.

Appraising the applicability and quality of economic evidence

The applicability and quality of economic evidence, including economic evidence derived from primary economic modelling conducted for the guideline, was assessed using the economic evaluations checklist specified in Developing NICE guidelines: the manual (NICE 2018a), Appendix H, for all studies that met the inclusion criteria.

The methodological assessment of economic studies considered in this guideline has been summarised in economic evidence profiles that were developed for each review question for which economic evidence was available. All studies that fully or partially met the applicability and quality criteria described in the methodology checklist were considered during the guideline development process.

Economic profiles of all economic studies that were considered during guideline development, including de novo economic analyses undertaken for this guideline, are provided in Appendix J of the respective Evidence Review Reports.

Inclusion and exclusion of health state utility studies

The titles and abstracts of papers identified through the searches were independently assessed for inclusion using predefined eligibility criteria defined in Table 11.

Table 11: Inclusion and exclusion criteria for the systematic review of health state utility values

Inclusion criteria
Only studies published from 2000 onwards were included in the review, so that evidence were relevant to current healthcare settings and preferences.
Studies were included provided that sufficient details regarding methods and results were available to enable the methodological quality of the study to be assessed, and provided that the study's data and results were extractable.
To be included, studies should report utility data for specific health states associated with acne through the care pathway.

Health-related quality of life should be rated directly by people with acne using a validated generic measure (such as EQ-5D, SF-6D, HUI-3) or a validated preference-based acne-specific measure, or a validated non-preference-based acne-specific measure that could be mapped onto a preference-based measure; alternatively, utility values could be derived by valuation of vignettes describing acne-related health states.
Valuation should be based on a choice-based method (i.e. time trade-off or standard gamble) and not on a visual analogue scale. Preferences could be derived from a sample of the general population or people with acne or their carers or health professionals. Preferences of the UK population were prioritised over preferences derived from non-UK populations.
Exclusion criteria
Poster presentations and abstracts in conference proceedings
Non-English language papers
Studies reporting an overall utility score for people with acne (and/or people without acne), who might have a mixture of acne-related health states or a range of symptom severity, were not considered.

Once the screening of titles and abstracts was complete, full versions of the selected papers were acquired for assessment.

Utility studies that met inclusion criteria and those that were excluded after full text was obtained are reported in the appendix J and appendix K, respectively, of evidence reports for areas in which economic modelling was undertaken.

Economic modelling

The aims of the economic input to the guideline were to inform the guideline committee of potential economic issues to ensure that recommendations represented a cost effective use of healthcare resources. Economic evaluations aim to integrate data on care benefits (ideally in terms of quality-adjusted life-years; QALYs) with the costs of different options. In addition, the economic input aimed to identify areas of high resource impact, as these need to be supported by robust evidence on cost effectiveness.

Areas for economic modelling were prioritised by the committee. The rationale for prioritising review questions for economic modelling was set out in an economic plan agreed between NICE, the committee, and members of the NGA technical team. Economic modelling was undertaken in areas with likely major resource implications, where the current extent of uncertainty over cost effectiveness was significant and economic analysis was expected to reduce this uncertainty. The following economic questions were selected as key issues that were addressed by economic modelling:

- Cost-effectiveness of treatments for people with mild to moderate acne. The methods and results of the de novo economic analysis are fully reported in appendix J of evidence review E1.
- Cost-effectiveness of treatments for people with moderate to severe acne. The methods and results of the de novo economic analysis are fully reported in appendix J of evidence review F1.
- Cost-effectiveness of interventions for the management of acne-related scarring. This question was not possible to model, due to lack of sufficient clinical evidence. Instead, a simple cost consequence analysis was undertaken, where intervention costs were assessed alongside intervention outcomes, in order to formulate

recommendations. The approach to this cost consequence analysis is described in evidence review M, under the 'Economic model' sub-heading.

When new economic analysis was not prioritised, the committee made a qualitative judgement regarding cost effectiveness by considering expected differences in resource and cost use between options, alongside clinical effectiveness evidence identified from the clinical evidence review.

Cost effectiveness criteria

NICE's report The NICE Principles sets out the principles that committees should consider when judging whether an intervention offers good value for money. In general, an intervention was considered to be cost effective if any of the following criteria applied (provided that the estimate was considered plausible):

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

The committee's considerations of cost effectiveness are discussed explicitly under the heading 'The committee's discussion of the evidence' under subheading 'Cost effectiveness and resource use' in the relevant evidence reviews.

Developing recommendations

Guideline recommendations

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

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

For further details refer to Developing NICE guidelines: the manual (NICE 2018a).

Research recommendations

When areas were identified for which evidence was lacking, the committee considered making recommendations for future research. For further details refer to Research recommendations process and methods guide (NICE 2015).

Validation process

This guideline was subject to a 6-week public consultation and feedback process. All comments received from registered stakeholders were responded to in writing and posted on the NICE website at publication. For further details refer to Developing NICE guidelines: the manual (NICE 2018a).

Updating the guideline

Following publication, NICE will undertake a surveillance review to determine whether the evidence base has progressed sufficiently to consider altering the guideline recommendations and warrant an update. For further details refer to Developing NICE guidelines: the manual (NICE 2018a).

Funding

The NGA was commissioned by NICE to develop this guideline.

References

Bradburn 2007

Bradburn, M. J., Deeks, J. J., Berlin, J. A., & Localio, A. R. Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events. Statistics in Medicine, 26, 53–77, 2007.

Caldwell 2005

Caldwell, D. M., Ades, A. E., & Higgins, J. P. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ, 331(7521), 897-900., 2005.

Dias 2010

Dias, S., Welton, N. J., Marinho, V. C. C., Salanti, G., Higgins, J. P. T., Ades, A. E. Estimation and adjustment of bias in randomised evidence by using Mixed Treatment Comparison Meta-analysis. Journal of the Royal Statistical Society (A), 173(3), 613-629, 2010.

Dias 2011

Dias, S., Welton, N. J., Sutton, A. J., & Ades, A. E. NICE DSU Technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials, 2011 (last updated 2016). http://www.nicedsu.org.uk

Dixon-Woods 2005

Dixon-Woods M, Agarwal S, Jones D et al. (2005) Synthesising qualitative and quantitative evidence: a review of possible methods. Journal of Health Services Research & Policy 10(1), 45–53

Hayden 2013

Jill A. Hayden, Danielle A. van der Windt, Jennifer L. Cartwright, Pierre Côté, Claire Bombardier. Assessing Bias in Studies of Prognostic Factors. Ann Intern Med. 2013;158:280–286. doi: 10.7326/0003-4819-158-4-201302190-00009

Higgins 2020

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). (2020) Cochrane Handbook for Systematic Reviews of Interventions Version 6.1.0 [updated September 2020] The Cochrane Collaboration. Available from https://www.training.cochrane.org/handbook (accessed 15 December 2020)

Lewin 2015

Lewin S, Glenton C, Munthe-Kaas H et al. (2015) Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS Med 12(10), e1001895

Lu 2004

Lu, G., & Ades, A. E. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med, 23(20), 3105-3124, 2004.

McGowan 2016

McGowan J, Sampson M, Salzwedel DM et al. (2016) PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. Journal of Clinical Epidemiology 75: 40–6

NICE 2015

National Institute for Health and Care Excellence (NICE) (2015) Research recommendations process and methods guide. Available from https://www.nice.org.uk/Media/Default/About/what-we-do/Research-and-development/Research-Recommendation-Process-and-Methods-Guide-2015.pdf (accessed 15 December 2020)

NICE 2018a

National Institute for Health and Care Excellence (NICE) (2014) Developing NICE guidelines: the manual (updated 2018). Available from https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview (accessed 15 December 2020)

NICE 2018b

National Institute for Health and Care Excellence (NICE) (2018) Policy on declaring and managing interests for NICE advisory committees. Available from https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-procedures/declaration-of-interests-policy.pdf (accessed 15 December 2020)

Phillippo 2019

Phillippo, D. M., Dias, S., Welton, N. J., Caldwell, D. M., Taske, N., Ades, A. E. Threshold Analysis as an Alternative to GRADE for Assessing Confidence in Guideline Recommendations Based on Network Meta-analyses. Ann Intern Med, 170(8), 538-46, 2019.

Santesso 2016

Santesso N, Carrasco-Labra A, Langendam M et al. (2016) Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. Journal of clinical epidemiology 74, 28-39

Van Valkenhoef 2016

van Valkenhoef, G., Dias, S., Ades, A.E., & Welton, N.J. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. Research Synthesis Methods, 7, 80-93, 2016

Welton 2009

Welton, N.J., Ades, A. E., Carlin, J. B., Altman, D. G., & Sterne J. A. C. Models for potentially biased evidence in meta-analysis using empirically based priors. Journal of the Royal Statistical Society (A), 172(1), 119-136, 2009.