National Institute for Health and Care Excellence

Final

Maternal and child nutrition

Methods

NICE guideline NG247

Supplement 1

January 2025

Final

Commissioned by the National Institute for Health and Care Excellence

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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

ISBN: 978-1-4731-6796-4

Contents

Development of the guideline

Remit

This guideline will update and amalgamate:

- the NICE guideline on maternal and child nutrition (PH11), and
- the recommendations on weight management during pregnancy in the NICE guideline on weight management before, during and after pregnancy (PH27).

(Note that the recommendations on weight management before and after pregnancy will be covered in the NICE guideline on overweight and obesity management).

What this guideline covers

Groups that will be covered

- Women during a single or multiple pregnancy (weight management and nutrition)
- Breastfeeding women (uptake of vitamins and maintaining breastfeeding)
- Preconception in relation to folic acid supplements only
- Babies and children from birth to 5 years and their parents and carers

Breastfeeding will only be covered from 8 weeks after birth. Feeding up to 8 weeks is covered in the NICE guideline on postnatal care.

We will give specific consideration to women living with underweight, overweight or obesity during pregnancy.

Settings that will be covered

All settings where publicly funded maternal and child nutrition assessment, advice and support is provided.

Key areas that will be covered in this update

We will look at evidence in the areas below when developing the guideline. We will consider making new recommendations or updating existing recommendations in these areas only. It may not be possible to make recommendations in all the areas.

1) Vitamin supplementation.
2) Weight management and healthy eating during pregnancy.
3) Breastfeeding and formula feeding.
4) Healthy eating behaviours in children up to 5 years.

This guideline will also link to any relevant recommendations on dietary advice, allergies and oral health in other NICE and government guidance.

What this guideline does not cover

Areas that will not be covered by this update

- Population-based screening programmes.

- Specialist dietary interventions for women and children following a specific diet for a medical condition.
- National maternal and child nutrition policies that are already covered by the Department of Health and Social Care (advised by SACN) and the Food Standards Agency (advised by the Committee on Toxicity), such as population-based dietary recommendations, national advice on food safety, the nutritional composition of infant formula and the fortification of foods.
- Interventions, information and support for breastfeeding and formula feeding of babies up to 8 weeks, as this is covered in the NICE guideline on postnatal care.
- Weight management for women before and after pregnancy, as these are covered by the update to the NICE guidelines on weight management.
- Weight management for children. Children aged over 2 years are covered by the update to the NICE guidelines on weight management. Weight management for children under 2 years will not be considered by thisguideline or the weight management guideline. It is felt that concerns in this area could be appropriately addressed by regular weight monitoring and by health professionals implementing existing advice on healthy eating behaviours in this population group.
- Care of preterm babies and low-birth-weight babies (defined by the World Health Organization as a birth weight less than 2,500 g).
- Complementary therapy

Methods

This guideline was developed using the methods described in the Developing NICE guidelines: the manual.

Declarations of interest were recorded according to the NICE conflicts of interest policy.

Developing the review questions and outcomes

The review questions developed for 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:

- population, intervention, comparator and outcome (PICO) for reviews of interventions
- 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 (PICo)

Full literature searches, critical appraisals 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 below.

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

Evidence review	Review question	Type of review
[A] High-dose folic acid supplementation before and during the first 12 weeks of pregnancy	Which groups of women should be advised to take high-dose folic acid supplements before and during the first 12 weeks of pregnancy?	Intervention
[B] Optimum folic acid supplementation dose before and during the first 12 weeks of pregnancy for women with a BMI ≥ 25 kg/m2 or more	What is the optimum dose of folic acid supplementation before and during the first 12 weeks of pregnancy for women with a BMI ≥ 25 kg/m2 or more?	Intervention
[C] Interventions to increase uptake of folic acid supplementation before and during the first 12 weeks of pregnancy	What interventions are effective to increase uptake of folic acid supplementation before and during the first 12 weeks of pregnancy?	Intervention
[D] Optimum vitamin D dose during pregnancy for women medically classified as overweight or obese	What dose of vitamin D is appropriate during pregnancy for women medically classified as overweight or obese?	Intervention

Evidence review	Review question	Type of review
[E] Interventions to increase uptake of vitamin supplements (including Healthy Start vitamins) in line with government advice	What interventions are effective to increase uptake of vitamin supplements (including Healthy Start vitamins) in line with government advice for pregnant women, breastfeeding women, babies and children up to 5 years?	Intervention
[F] Healthy and appropriate weight change during pregnancy	What gestational weight change is healthy and appropriate during pregnancy?	Prognostic
[G] Interventions for helping to achieve healthy and appropriate weight change during pregnancy	What are the most effective and cost-effective interventions for helping women to achieve healthy and appropriate weight change during pregnancy?	Intervention
[H] Healthy lifestyle interventions for those with gestational diabetes	What are the most effective and cost-effective healthy lifestyle interventions for women with gestational diabetes?	Intervention
[I] Interventions to increase uptake of healthy eating and drinking advice during pregnancy	What interventions are effective to increase uptake of healthy eating and drinking advice during pregnancy in line with government advice?	Intervention
[J] Approaches and interventions for maintaining breastfeeding beyond 8 weeks after birth	What approaches and interventions are effective in maintaining breastfeeding after 8 weeks?	Intervention
[K] Facilitators and barriers for maintaining breastfeeding beyond 8 weeks after birth	What do parents perceive to be facilitators and barriers for maintaining breastfeeding after 8 weeks?	Qualitative
[L] Facilitators and barriers to follow existing government advice on safe and appropriate formula feeding	What are the facilitators and barriers for parents to follow existing government advice on safe and appropriate formula feeding?	Qualitative
[M] Facilitators and barriers to continue breastfeeding when returning to work or study	What are the facilitators and barriers to help women returning to work and study to continue breastfeeding?	Qualitative
[N] Interventions to promote appropriate and timely introduction to solids (complementary feeding) for babies from 6 to 12 months	What interventions are effective to promote appropriate and timely introduction to solids (complementary feeding) for babies from 6 to 12 months (in line with government advice)?	Intervention
[O] Interventions to promote healthy eating and drinking practices, including	What interventions are effective to promote healthy eating and drinking practices, including complementary	Intervention

Evidence review	Review question	Type of review
complementary feeding, in children from 12 months to 5 years	feeding, in children from 12 months to 5 years (in line with government advice)?	
[P] Facilitators and barriers to increase the uptake of government advice on folic acid and vitamin supplements	What are the barriers and facilitators to increasing the uptake of government advice for women and families with children up to five years in the following areas: • folic acid supplements (including before pregnancy) • vitamin supplements (including Healthy Start vitamins)?	Qualitative
[Q] Facilitators and barriers to increase the uptake of government advice on healthy eating and drinking in pregnancy	What are the barriers and facilitators to increasing the uptake of government advice for women and families with children up to five years in the following areas: • healthy eating and drinking in pregnant women?	Qualitative
[R] Facilitators and barriers to increase the uptake of government advice on appropriate and timely introduction to solids and healthy eating and drinking in children	What are the barriers and facilitators to increasing the uptake of government advice for women and families with children up to five years in the following areas: • appropriate and timely introduction to solids (complementary feeding) for babies from 6 to 12 months • healthy eating and drinking in children from 12 months to 5 years?	Qualitative

The COMET database was searched for core outcome sets relevant to this guideline. A core outcome set for maternal and fetal/childhood outcomes (set by Mehra 2012 and Farpour-Lambert 2018) were used in the evidence reviews. Additional outcomes on healthy eating and drinking during pregnancy and in children were chosen based on committee discussions.

Additional information related to development of the guideline is contained in Supplement 2 NICE technical team list.

Searching for evidence

Scoping search

During the scoping phase, searches were conducted for previous guidelines, economic evaluations, health technology assessments, systematic reviews, randomised controlled trials, observational studies and qualitative research.

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 the searches were conducted in the following databases: Medline ALL and Embase.

For review questions related to interventions the following databases were also searched: Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR), Epistemonikos and Cinahl. For qualitative review questions Emcare and PsycINFO were also searched.

Searches were run once for all reviews during development. Searches for the following questions were updated in December 2023, 13 weeks in advance of the final committee meeting.

- [A] High dose folic acid supplementation
- [B] Folic acid supplementation for women with a BMI ≥ 25 kg/m2 or more
- [D] Vitamin D dose during pregnancy for women medically classified as overweight or obese
- [E] Interventions to increase uptake of vitamin supplements (including Healthy Start vitamins) in line with government advice

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. Databases were searched using subject headings, free-text terms and, where appropriate, an economic evaluations search filter.

Searches using the search strategies derived from the review questions, combined with a search filter for economic evaluations, were conducted in Medline ALL, Embase, INAHTA (International HTA Database) and CRD HTA. Where possible, searches were limited to studies published in English. Limits to exclude animal studies, letters, editorials, news were applied where possible.

As with the general literature searches, the economic literature searches were run once for all reviews during development. Searches for the following questions were updated in December 2023, 13 weeks in advance of the final committee meeting.

- [A] High dose folic acid supplementation
- [B] Folic acid supplementation for women with a BMI ≥ 25 kg/m2 or more
- [D] Vitamin D dose during pregnancy for women medically classified as overweight or obese
- [E] Interventions to increase uptake of vitamin supplements (including Healthy Start vitamins) in line with government advice

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

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).

Reviewing research 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 D of each evidence review).
- Included studies were critically appraised using an appropriate checklist as specified in Developing NICE guidelines: the manual. Further detail on appraisal of the evidence is provided below.
- Summaries of quantitative evidence by outcome and qualitative evidence by theme were presented in the corresponding evidence review and discussed by the committee.

Review questions selected as high priorities for economic analysis (and those selected as medium priorities and where economic analysis could influence recommendations) and complex review questions 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. The review protocol for each question specifies whether dual screening and study selection was undertaken for that particular question. Drafts of all evidence reviews were quality assured 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 insufficient evidence from RCTs to inform guideline decision making, non-randomised studies (NRS) were considered for inclusion. Sufficiency was judged

taking into account the number, quality and sample size of RCTs, as well as outcomes reported and availability of data from subgroups of interest. When NRS were considered for inclusion, priority was given to controlled studies, with separate control groups that were not allocated on the basis of the outcome, that adjusted for relevant confounders or matched participants on important confounding domains.

For prognostic reviews, prospective and retrospective cohort studies 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.

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 J 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 studies

Pairwise meta-analysis

Meta-analysis to pool results from comparative intervention studies was conducted where possible using Cochrane Review Manager (RevMan5) software.

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. 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. Where a study reported multiple adjusted estimates for the same outcome, the one that minimised the risk of bias due to confounding was chosen.

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 or ratings based on sample size cut-offs were considered instead.

For some reviews, evidence was either stratified from the outset or separated into subgroups when heterogeneity was encountered. The stratifications and potential subgroups were pre-defined at the protocol stage (see the protocols for each review for further detail). Where evidence was stratified or sub-grouped the committee considered on a case-by-case basis if separate recommendations should be made for distinct groups. Separate recommendations may be made where there is evidence of a differential effect of interventions in distinct groups. If there is a lack of evidence in one group, the committee considered, based on their experience, whether it was reasonable to extrapolate and assume the interventions will have similar effects in that group compared with others.

Where applicable, data from RCTs and NRS, or from NRS with substantially different designs (i.e., cohort studies and case-control studies), that were theoretically possible to pool were entered into RevMan5 as subgroups based on study design. This was to take into account the likelihood of increased heterogeneity from studies with different design features and different approaches to appraising the quality of evidence based on study design (see appraising the quality of evidence: intervention studies below).

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

Meta-regression

Meta-regression analysis was considered appropriate to assess the effectiveness of education, advice or support interventions aimed to maintain breastfeeding beyond 8 weeks after birth, covered in evidence review J. Meta-regression is used in meta-analysis to simultaneously investigate the impact of moderator variables on study effect size. In this case meta-regression was considered appropriate because there was a large volume of included studies (n=69) each with different intervention characteristics (or 'moderator variables'), for example where the intervention was delivered, how long it lasted for, how the intervention was delivered and how often.

For the purpose of this meta-regression analysis, each study was categorised using the following variables.

- Number of contact visits: 0, 1, 2-3, 4-8 and 9+.
- How delivered: face to face on an individual basis, face to face in a group, remote, self-help.
- Duration of contact: contact with the intervention lasted less than 8 weeks, contact with the intervention lasted more than 8 weeks.
- Where the intervention was delivered: at home, in a healthcare setting, combination of both home and healthcare setting.

The following analyses were conducted for each outcome (i.e. any breastfeeding at 6-12 weeks, exclusive breastfeeding at 6-12 weeks, any breastfeeding at 16-26 weeks, exclusive breastfeeding at 16-26 weeks).

- How delivered

  - Face to face as an individual versus standard care
  - Remote versus standard care
  - Self-help versus standard care
- Number of contacts
  - 0-1 versus standard care
  - 2-3 versus standard care
  - 4-8 versus standard care
  - 9+ versus standard care
- Duration of contact
  - Less than 8 weeks versus standard care
  - More than 8 weeks versus standard care
- Where delivered
  - Healthcare setting versus standard care
  - Home setting versus standard care
  - Both healthcare and home setting versus standard care

Individual models were first run for each of the variable categories (number of contacts, how delivered, duration of contact and where the intervention was delivered). We attempted to run a final 'combined' model, ideally incorporating all variables in one analysis. However, there was significant collinearity between the variables, which did not allow the model to converge. To avoid this, a number of variables and/or categories within variables needed to be omitted or merged – this considerably reduced the information provided by the combined model and increased the uncertainty around the resulting study effects, so it was decided not to consider an analysis using the combined model.

Meta-regression was implemented in WinBUGS 1.4.3 (Spiegelhalter 2003). A sample WinBUGS code for the analysis of any breastfeeding at 16 to 26 weeks, including the variables how the intervention was delivered, the number of contacts for the intervention and where the intervention was delivered is given in evidence review J, appendix M. Other analyses used the same substantive code as the one provided, modified to include the relevant predictor variables for the model under consideration.

See evidence review J for further details of the meta-regression methods and results.

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. Meta-analysis using the same methods as for intervention reviews outlined above was performed where possible (for example, if there were at least 2 studies reporting the same risk factor and in populations with the same/similar characteristics) and where there was no significant variation between studies or very serious heterogeneity. For those where meta-analysis could not be performed, the results for each individual study have been reported in the review.

Data synthesis for qualitative reviews

Where possible, a meta-synthesis was conducted to combine evidence from more than one study into a theme or sub-theme. 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.

The technical team were guided in their data extraction, synthesis and formulation of review findings, or themes, by a framework of phenomena developed by the guideline committee. This framework consisted of the themes that the committee anticipated would be covered by the included studies and these were set out a priori in the corresponding review protocol. As well as guiding the data extraction and synthesis, the framework also underpinned the approach referred to in the protocol as 'thematic saturation'. Essentially, data or themes from included studies would not be extracted if they contributed to review findings which were judged to be 'adequate' and 'coherent' following assessment using the GRADE-CERQual approach; that is, they were not downgraded for either domain. Themes identified from the included studies, which were not set out in the protocol but which were considered relevant to answering the review question, were also extracted and the same approach to 'thematic saturation' would have been applied. Thematic saturation was not reached for any themes in any of the qualitative components of the reviews in this guideline. Therefore, all relevant data from all included qualitative studies were extracted and analysed.

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 3. Criteria considered in the rating of these elements are discussed below. Each element was graded using the quality ratings summarised in Table 4. 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 5.

The initial quality rating was based on the study design: RCTs and NRS assessed by ROBINS-I start as 'high' quality evidence, other non-randomised studies start as 'low' quality evidence. The rating was then modified according to the assessment of each quality element (Table 3). 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 (RoB 2; see Appendix H in Developing NICE guidelines: the manual).

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

- selection bias
- performance bias
- attrition bias
- detection bias
- reporting bias.

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 can be found in Section 8 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011).

For systematic reviews the ROBIS checklist was used (see Appendix H in Developing NICE guidelines: the manual).

For non-randomised controlled studies, cohort studies, uncontrolled before after studies or historical controlled studies the ROBINS-I checklist was used (see Appendix H in Developing NICE guidelines: the manual).

For controlled before after studies, the EPOC risk of bias tool was used (see Appendix H in Developing NICE guidelines: the manual).

For cross sectional studies-, the JBI Checklist for Analytical Cross Sectional Studies was used (see Appendix H in Developing NICE guidelines: the manual).

Wang 2021 checklist was used for assessing the methodological quality of IPD meta-analysis (Wang 2021) (see Appendix H in Developing NICE guidelines: the manual).

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 serious 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 no plausible explanation for the serious or very serious heterogeneity could be found, the quality of the evidence was downgraded in GRADE for inconsistency and the meta-analysis was re-run using the Der-Simonian and Laird method with a random effects model.

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.

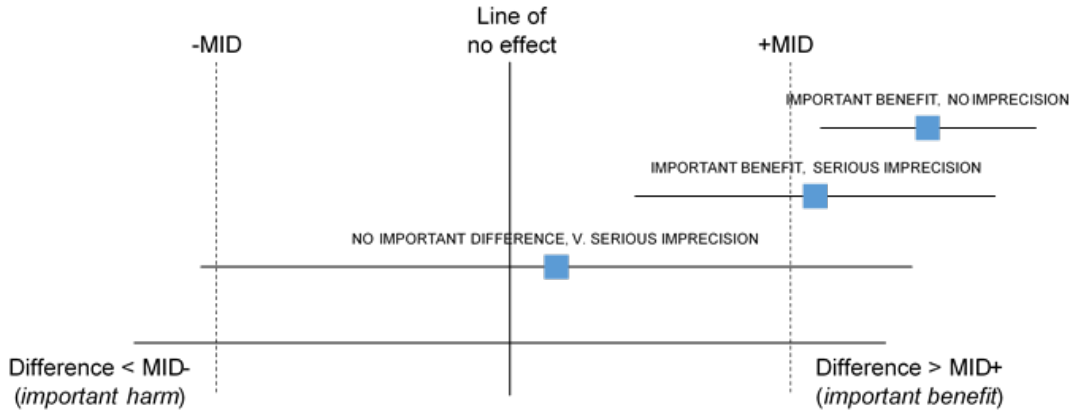
When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

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.

The committee agreed that there were a number of outcomes, namely caesarean birth, hypertensive disorders of pregnancy, gestational diabetes, small for gestational age or large for gestational age, that were sufficiently serious that any statistically

significant difference would be considered clinically important. In such cases, imprecision was assessed based on the total number of events (>300 events: no imprecision;150-300 events: serious imprecision; <150 events: very serious imprecision) for dichotomous outcomes and total sample size for continuous outcomes (>400 people: no imprecision; 200-400 people: serious imprecision; <200 people: very serious imprecision). The committee used these numbers based on commonly used optimal information size thresholds.

For the remaining outcomes, 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 200 and 400 as cut-offs for very serious and serious imprecision, respectively.

For continuous outcomes GRADE 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 at baseline). Where results were reported as medians, imprecision was assessed based on sample size using 200 and 400 as cut-offs for very serious and serious imprecision, respectively.

MIDs, the line of no effect, and both 95% and 90% confidence intervals (CIs) were used to assess whether there were important differences in outcomes between groups. Outcomes were considered to have an important benefit/harm, possible important benefit/harm, no evidence of an important difference, or no important difference using the following approach:

- Where the point estimate (PE) was greater than the upper MID and the 95% CI did not cross line of no effect, an intervention was described as having an important benefit
- Where the PE was greater than the upper MID and the 95% CI crossed the line of no effect, but the 90% CI did not, an intervention was described as having a possible important benefit
- Where the PE was greater than the upper MID or lower than the lower MID, and the 90% CI crossed the line of no effect, the result was described as no evidence of an important difference
- Where the PE was between two MIDs, the result was described as no important difference
- Where the PE was lower than the lower MID and the 95% CI crossed the line of no effect, but the 90% CI did not, an intervention was described as having a possible important harm
- Where the PE was lower than the lower MID and the 95% CI did not cross line of no effect, an intervention was described as having an important harm.

This approach was used for all evidence reviews which informed decision making on the guideline. Please note that the above descriptions were based on positive outcomes (where high values indicate better outcomes or events are positive). If the outcomes were negative (where high values indicate worse outcomes or events are negative) then whether an intervention is considered to have an important benefit or important harm would be switched (for example, where the PE is greater than the upper MID and the 95% CI do not cross the line of no effect, an intervention would be described as having an important harm; where the PE is lower than the lower MID and the 95% CI do not cross line of no effect, an intervention would be described as having an important benefit).

Assessing publication bias in intervention reviews

We did not assess publication bias for intervention reviews in this guideline.

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.

The evidence for each outcome in the prognostic reviews was examined separately for the quality elements listed and defined in Table 6. The criteria considered in the rating of these elements are discussed below. Each element was graded using the quality levels summarised in Table 4. 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 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 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). 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 serious heterogeneity, and more than 80% indicating very serious heterogeneity. When serious 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, data were not pooled.

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 to assess imprecision. Clinical importance was assessed by the association between the risk factor and the outcome, and the committee agreed that any statistically significant association between the risk factors and outcomes was clinically important.

Qualitative studies

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 2018) 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 8. Each element was graded using the levels of concern summarised in Table 9.

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 10. 'Confidence' in this context refers to the extent to which the review finding is a reasonable representation of the phenomenon of interest set out in the protocol. Similar to other types of evidence all review findings start off with 'high confidence' and are rated down by one or more levels if there are concerns about any of the individual CERQual components. In line with advice from the CERQual developers, the overall assessment does not involve numerical scoring for each component but in order to ensure consistency across and between guidelines, the NGA established some guiding principles for overall ratings. For example, a review finding would not be downgraded (and therefore would be assessed with 'high' confidence) if at least 2 of the individual components were rated as 'no or very minor; and none of the components were rated as having moderate or serious concerns.

At the other extreme, a review finding would be downgraded 3 times (to 'very low') if at least 2 components had serious concerns or 3 had moderate concerns (as long as the 4th component was rated 'serious') or if all components had moderate concerns. A basic principle was that if any components had any serious concerns, then overall confidence in the review finding would be downgraded at least twice, to low. Transparency about overall judgements is provided in the CERQual tables, with explanations for downgrading given in table footnotes.

Table 6: Adaptation of GRADE quality elements for qualitative reviews

Quality element	Description
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 our confidence that the review findings reflect the phenomena of interest. 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 context of the studies 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. If the data from the underlying studies are ambiguous or contradict the review finding this would reduce our confidence in the finding.
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. Judgements are not based on the number of studies but do take account of the quantity and also richness of data underpinning a finding. The more complex the finding, the more detailed the supporting data need to be. For simple findings, relatively superficial data would be considered adequate to explain and explore the phenomenon being described.

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). Overall methodological limitations were derived by assessing the methodological limitations across the 6 domains summarised in Table 11.

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 health or social care 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 it does take account of the quantity of data supporting a review finding (for instance whether sufficient quotations or observations were provided to underpin the findings) and in particular the degree of ‘richness’ of supporting data. Concerns about richness arise when insufficient details are provided by the data to enable an understanding of the phenomenon being described. Generally, if a review finding is fairly simple then relatively superficial data will be needed to understand it. Data underpinning a more complex finding would need to

offer greater detail, allowing for interpretation and exploration of the phenomenon being described. Therefore, in assessing adequacy our downgrading involved weighing up the complexity of the review finding against the explanatory contribution of the supporting data.

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 evidence were conducted in all areas covered in the guideline, as relevant. Reviews of economic evidence were not relevant for questions addressed by reviews of qualitative evidence. Titles and abstracts of articles identified through the economic literature searches were independently assessed for inclusion using the predefined eligibility criteria listed in Table 13.

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

Inclusion criteria
For each review question, selection criteria regarding the study population and the interventions or conditions assessed were identical to those described in the respective effectiveness review protocol.
Only studies from the 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 2002 onwards were included in the review. This date restriction was imposed so that retrieved economic evidence was relevant to current healthcare settings and costs.
Only studies that reported sufficient details regarding methods and results, to enable the methodological quality of the study to be assessed were included, provided also that the study's data and results were extractable.
Full economic evaluations that compared 2 or more relevant options and considered both costs and consequences as well as costing analyses that compared only costs between 2 or more interventions.
Clinical effectiveness data utilised in the analysis should have been derived from a literature review, a clinical trial, a prospective or retrospective cohort study, or a study with a before-and-after design.
Studies should be reporting separately costs for each option assessed, from a healthcare perspective.
Exclusion criteria
Poster presentations and abstracts in conference proceedings.
Non-English language papers.
Non-comparative studies.
Studies not reporting intervention costs.
Studies reporting exclusively intervention and/or implementation costs without any assessment of benefits or cost-savings.
Studies that adopted a non-healthcare perspective and did not consider healthcare costs.

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

Details of economic evidence study selection, lists of excluded studies, economic evidence tables, the results of quality assessment of economic evidence (see below) and health economic evidence profiles are presented in respective evidence reviews.

Appraising the 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, 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.

Inclusion and exclusion of health state utility studies

Literature on the health-related quality of life of populations covered in this guideline was systematically searched to identify studies reporting appropriate utility scores that could be utilised in a primary economic modelling. 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 from Organisation for Economic Co-operation and Development member countries were included, as the aim of the review was to identify utility data transferable to the UK context.
Studies should report utility data for health states associated with the populations covered in the guideline.
Studies should report health-related quality of life ratings made using a validated generic or harmful gambling-specific preference-based measure directly or via mapping from another validated non-preference-based measure. Utility values should have been elicited from the general population using a choice-based method, such as time trade-off (TTO) or standard gamble (SG).
Exclusion criteria
Poster presentations and abstracts in conference proceedings
Non-English language papers

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 listed in evidence review F, which included economic modelling.

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 healthcare 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; these are recommendations which (while cost-effective) might have a large impact on Clinical Commissioning Group or Trust finances and so need special attention.

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 the guideline technical team, the committee, and the NICE quality assurance 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 to be addressed by economic modelling:

- Cost-effectiveness of interventions aimed to increase uptake of folic acid before and during the first 12 weeks of pregnancy, focusing on health technologies. No economic modelling was carried out for this question, due to the limited amount and quality of the clinical evidence, which did not allow for a robust model to be developed or for recommendations on specific interventions to be made.
- Cost-effectiveness of interventions aimed to increase uptake of vitamin supplements (including Healthy Start vitamins) in line with government advice for pregnant women, breastfeeding women, babies and children up to 5 years, focusing on health technologies. No economic modelling was carried out for this question, due to the limited amount and quality of the clinical evidence, which did not allow for a robust model to be developed or for recommendations on specific interventions to be made.
- Cost-effectiveness of interventions that help women to achieve healthy and appropriate weight gain during pregnancy (for example, dietary interventions, regular weighing, physical activity). No economic modelling was carried out for this question, as the clinical evidence showed very small benefits and the committee did not wish to make recommendations on specific interventions; therefore, development of an economic model was not deemed useful.
- Cost-effectiveness of education, advice or support interventions aimed to maintain breastfeeding beyond 8 weeks after birth.

The methods and results of the de novo economic analysis carried out for the guideline are reported in Appendix I of the relevant evidence review. Where new economic analysis was not prioritised and no economic evidence was identified, 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 Our 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, qualitative 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.

Research recommendations

When areas were identified for which evidence was lacking, the committee considered making recommendations for future research. For further details refer to Developing NICE guidelines: the manual and NICE's Research recommendations process and methods guide.

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.

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.

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.

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

Farpour-Lambert Nathalie 2018

Farpour-Lambert Nathalie J. , Ells Louisa J. , Martinez de Tejada Begoña , Scott Courtney. Obesity and Weight Gain in Pregnancy and Postpartum: an Evidence Review of Lifestyle Interventions to Inform Maternal and Child Health Policies. Frontiers in Endocrinology. Volume 9. 2018. https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2018.00546 (accessed on 8th March 2024)

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 2011

Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated 2019] The Cochrane Collaboration. Available from www.handbook.cochrane.org (accessed on 1st Feb 2024)

Lewin 2018

Lewin S, Booth A, Glenton C, Munthe-Kaas H et al. (2018) Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci. 2018 Jan 25;13 (Suppl1): 2

Mehra 2012

Mehra H, Thangaratinam S. Prioritisation of outcomes in the evaluation of weight management interventions in pregnancy: a DELPHI survey. Archives of Disease in Childhood-Fetal and Neonatal Edition. 2012 Apr 1;97(Suppl 1):A38

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

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

Wang 2021

Huan Wang, Yancong Chen, Yali Lin, Julius Abesig, Irene XY Wu, Wilson Tam (2021) The methodological quality of individual participant data meta-analysis on intervention effects: systematic review. The methodological quality of individual participant data meta-analysis on intervention effects: systematic review (bmj.com) (accessed on 2nd Feb 2024)