National Institute for Health and Care Excellence

Final

Social, emotional and mental wellbeing in primary and secondary education

NICE guideline: methods

NICE guideline NG223

Methods

July 2022

FINAL

These evidence reviews were developed by developed by the Public Health Guidelines team

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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: 978-1-4731-4634-1

Contents

Development of the guideline

Remit

This guideline will update and replace NICE guidelines on social and emotional wellbeing in primary education (PH12) and social and emotional wellbeing in secondary education (PH20). For further detail of what the guideline covers please see the final scope document.

What this guideline covers

Whole-school approaches

1. Integrated approaches that include and go beyond teaching and learning in the classroom to all aspects of the life of a school including culture, ethos and environment, as well as partnerships with parents or carers and families, outside agencies, and the wider community.
2. Identifying vulnerable children and young people as part of the whole school approach.

Universal approaches

3. Curriculum content and classroom-based interventions focused on social, emotional and mental wellbeing. This includes lessons on resilience, self-esteem, coping skills (such as dealing with bereavement or adverse childhood events), mental health awareness, managing social relationships (to avoid bullying, including online bullying) and the appropriate and safe use of the internet and social media.

Targeted approaches

4. Targeted social or emotional support such as individual or small group interventions for areas such as self-esteem, resilience or coping skills for children and young people who need extra support in developing social and emotional skills.
5. Targeted mental health support such as individual or small group interventions for children and young people at risk of depression, anxiety or stress.

Transition

6. Support during periods of transition (for example developmental transitions such as puberty, life transitions such as family break-ups or bereavement, and educational transitions such as moving from primary to secondary school).

What this guideline does not cover

1. Interventions aimed at treating depression, anxiety or other mental health diagnoses.
2. Management of disruptive or violent behaviour.
3. Strategies focused on preventing self-harm or suicide.

Methods

This guideline was developed in accordance with the process set out in 'Developing NICE guidelines: the manual (2020)'. Where the guidelines manual does not provide advice, additional methods are described below.

Developing the review questions and outcomes

The 18 review questions developed for this guideline were based on the key areas identified in the guideline scope. Review questions were developed by the NICE Public Health Internal Guideline Development (PHIGD) team and refined, validated and signed off by the Public Health Advisory Committee (PHAC) and NICE quality assurance team.

The review questions were based on the PICO[S] framework - Population, Intervention, Comparator and Outcome [and Study type] for reviews of interventions

Full literature searches, critical appraisals and evidence reviews were completed for all review questions.

Details of these elements are found in the review protocols for each review (see Appendix A of each relevant review). Where protocol deviations have been made, these will be reported in the Methods section of the individual review.

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

Evidence review	Review questions	Type of review
A	1.1 What principles or combination of principles of whole-school approaches to promote social, emotional and mental wellbeing are effective and cost-effective? a) in children in primary education b) in children and young people in secondary and further education 1.2 Are whole-school approach interventions to promote the social, emotional and mental wellbeing of children and young people acceptable to • children and young people, • their parents or carers • the teacher and professionals delivering the interventions 1.3 What are the barriers and facilitators to using the whole-school approach to promote social, emotional and mental wellbeing in children and young people?	Convergent segregated mixed methods review

Evidence review	Review questions	Type of review
B	3.1a What universal classroom-based interventions to promote social, emotional and mental wellbeing in children in primary education are effective and cost effective? 3.1b What universal classroom-based interventions to promote social, emotional and mental wellbeing in children and young people in secondary and further education are effective and cost effective?	Quantitative element of a convergent segregated mixed methods review
C	3.2 Are universal classroom-based interventions acceptable to the children and young people receiving them, their parents or carers and to those delivering them? 3.3 What are the barriers and facilitators to using universal classroom-based interventions to promote social, emotional and mental wellbeing in children and young people?	Qualitative element of a convergent segregated mixed methods review (including mixed methods and committee discussion)
D	2.1 What are the risk factors associated with social, emotional and mental wellbeing?	Predictive association review
E	2.2 What are the barriers and facilitators to identifying children and young people at risk of poor social, emotional and mental wellbeing?	Qualitative evidence synthesis*
F	2.3 What is the usefulness (effectiveness and acceptability) of assessment tools to assess need for additional SEMW support in children and young people who have been identified as having poor social, emotional and mental wellbeing using 'soft intelligence' for example behaviours, school attendance, drop off in engagement?	Prognostic test accuracy
G	4.1a What is the effectiveness and cost-effectiveness of targeted interventions that aim to promote social and emotional support in children in primary education? 4.1 b What is the effectiveness and cost-effectiveness of targeted interventions that aim to promote social and emotional support in children and young people in secondary and further education? 4.2 Are targeted approaches to promote social, emotional and mental wellbeing acceptable to: • Children and young people receiving them • Teachers/practitioners delivering the interventions	Convergent segregated mixed methods review

Evidence review	Review questions	Type of review
	• Parents/Carers of children and young people receiving the interventions 4.3 What are the barriers and facilitators to using targeted approaches to promote social, emotional and mental wellbeing in children and young people?	
H	5.1a What is the effectiveness and cost-effectiveness of targeted mental health support approaches for children in primary education? 5.1b What is the effectiveness and cost-effectiveness of targeted mental health support approaches for children and young people in secondary and further education? 5.2 Are targeted mental health support approaches acceptable to • Children and young people receiving them • Teachers/practitioners delivering the interventions • Parents/Carers of children and young people receiving the interventions 5.3 What are the barriers and facilitators to using targeted mental health support?	Convergent segregated mixed methods review
I	6.1 What are effective and cost-effective interventions to support the social, emotional and mental wellbeing of children during periods of transition (such as between schools, life stages or due to traumatic events)? 6.2 Are interventions to support the social, emotional and mental wellbeing of children and young people during periods of transition (such as between schools, life stages or due to traumatic events) acceptable to: • Children and young people • Teachers/practitioners delivering the interventions • Parents/Carers of children and young people receiving the interventions • Schools/teachers dealing with the consequences of transition e.g. secondary schools dealing with a child's transition from primary to secondary school?	convergent segregated mixed methods review

Evidence review	Review questions	Type of review
	6.3 What are the barriers and facilitators to transition based interventions to promote social, emotional and mental wellbeing in children and young people?	

* Mixed methods review in protocol but no quantitative prognostic evidence identified

Reviewing research evidence

Review protocols

Review protocols were developed with the guideline committee to outline the inclusion and exclusion criteria used to select studies for each evidence review. Where possible, review protocols were prospectively registered in the PROSPERO register of systematic reviews. Protocols are reproduced in each evidence review along with the PROPSERO registration number if the protocol was registered.

Searching for evidence

Evidence was searched for each review question using the methods specified in the 2020 NICE guidelines manual. Brief details of search strategies can be found in the appendices of each individual review. Full details of search strategies, databases searched and numbers of studies identified can be found in the search chapter on the guideline webpage.

Selecting studies for inclusion

All references identified by the literature searches and from other sources (for example, previous versions of the guideline or studies identified by committee members) were uploaded into EPPI reviewer software (version 5) and de-duplicated. Titles and abstracts were assessed for possible inclusion using the criteria specified in the review protocol. 10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer.

All of the evidence reviews made use of the priority screening functionality within the EPPI-reviewer software. This functionality uses a machine learning algorithm (specifically, an SGD classifier) to take information on features (1, 2 and 3 word blocks) in the titles and abstract of papers marked as being ‘includes’ or ‘excludes’ during the title and abstract screening process, and re-orders the remaining records from most likely to least likely to be an include, based on that algorithm. This re-ordering of the remaining records occurs every time 25 additional records have been screened. Research is currently ongoing as to what are the appropriate thresholds where reviewing of abstracts can be stopped, assuming a defined threshold for the proportion of relevant papers it is acceptable to miss on primary screening. As a conservative approach until that research has been completed, the following rules were adopted during the production of this guideline:

- In every review, at least 50% of the identified abstracts (or 1,000 records, if that is a greater number) were always screened.

- After this point, screening was only terminated if a pre-specified threshold was met for a number of abstracts being screened without a single new include being identified. This threshold was set according to the expected proportion of includes in the review (with reviews with a lower proportion of includes needing a higher number of papers without an identified study to justify termination) and was always a minimum of 250.

As an additional check to ensure this approach did not miss relevant studies, systematic reviews (or qualitative evidence syntheses in the case of reviews of qualitative studies) were included in the review protocol and search strategy for all review questions. Relevant systematic reviews or qualitative evidence syntheses were used to identify any papers not found through the primary search. Committee members were also consulted to identify studies that were missed. If additional studies were found that were erroneously excluded during the priority screening process, the full database was subsequently screened.

The decision to use priority screening was taken on a case-by-case basis by the reviewing team depending on the perceived likelihood that stopping criteria would be met, based on the size of the database, heterogeneity of studies included in the review and predicted number of includes. If it was thought that stopping criteria were unlikely to be met, priority screening was not used, and the full database was screened.

The full text of potentially eligible studies was retrieved and assessed according to the criteria specified in the review protocol. A standardised form was used to extract data from included studies into the EPPI reviewer software. Study investigators were contacted for missing data when time and resources allowed (when this occurred, this was noted in the evidence review and relevant data was included).

Incorporating published evidence syntheses

For all review questions where a literature search was undertaken looking for a particular study design, published evidence syntheses (quantitative systematic reviews or qualitative evidence syntheses) containing studies of that design were also included. All included studies from those syntheses were screened to identify any additional relevant primary studies not found as part of the initial search. Evidence syntheses that were used solely as a source of primary studies were not formally included in the evidence review (as they did not provide additional data) and were not quality assessed.

Methods of combining evidence

Data synthesis for intervention studies

Where possible, meta-analyses were conducted to combine the results of quantitative studies for each outcome.

Pairwise meta-analysis

Pairwise meta-analyses were performed in Cochrane Review Manager V5.3 where possible. Meta-analyses that could not be conducted in Cochrane Review Manager were carried out in R version 3.3.4. using the package ‘metafor’. A pooled relative risk was calculated for dichotomous outcomes (using the Mantel–Haenszel method) reporting numbers of people having an event. Both relative and absolute risks were presented, with absolute risks calculated by applying the relative risk to the risk in the

comparator arm of the meta-analysis (calculated as the total number events in the comparator arms of studies in the meta-analysis divided by the total number of participants in the comparator arms of studies in the meta-analysis).

A pooled mean difference was calculated for continuous outcomes (using the inverse variance method) when the same scale was used to measure an outcome across different studies. Where different studies presented continuous data measuring the same outcome but using different numerical scales (e.g. a 0-10 and a 0-100 visual analogue scale), these outcomes were all converted to the same scale before meta-analysis was conducted on the mean differences. Where outcomes measured the same underlying construct but used different instruments/metrics, data were analysed using standardised mean differences (SMDs, Hedges' g).

For continuous outcomes analysed as mean differences, change from baseline values were used in the meta-analysis if they were accompanied by a measure of spread (for example standard deviation). Where change from baseline (accompanied by a measure of spread) were not reported, the corresponding values at the timepoint of interest were used. If only a subset of trials reported change from baseline data, final timepoint values were combined with change from baseline values to produce summary estimates of effect. For continuous outcomes analysed as standardised mean differences this was not possible. In this case, if all studies reported final timepoint data, this was used in the analysis. If some studies only reported data as a change from baseline, analysis was done on these data, and for studies where only baseline and final time point values were available, change from baseline standard deviations were estimated, assuming a correlation coefficient derived from studies reporting both baseline and endpoint data, or if no such studies were available, assuming a correlation of 0.5 as a conservative estimate (Follman et al., 1992; Fu et al., 2013).. In cases where SMDs were used they were back converted to a single scale to aid interpretation by the committee where possible.

Random effects models were fitted when there was significant between-study heterogeneity in methodology, population, intervention or comparator was identified by the reviewer in advance of data analysis. This decision was made and recorded before any data analysis was undertaken.

For all other syntheses, fixed- and random-effects models were fitted, with the presented analysis dependent on the degree of heterogeneity in the assembled evidence. Fixed-effects models were the preferred choice to report, but in situations where the assumption of a shared mean for fixed-effects model were clearly not met, even after appropriate pre-specified subgroup analyses were conducted, random-effects results are presented. Fixed-effects models were deemed to be inappropriate if there was significant statistical heterogeneity in the meta-analysis, defined as $I^2 \geq 50\%$.

However, in cases where the results from individual pre-specified subgroup analyses were less heterogeneous (with $I^2 < 50\%$) the results from these subgroups were reported using fixed effects models. This may have led to situations where pooled results were reported from random-effects models and subgroup results were reported from fixed-effects models.

Where sufficient studies were available, meta-regression was considered to explore the effect of study level covariates.

Data synthesis for predictive accuracy data

For the purpose of this guideline predictive accuracy data are classified as any data in which an index feature - be it a symptom, a risk factor, a test result or the output of some algorithm that combines many such features - is observed in some people who develop a condition or outcome of interest at some time after the observation of the index feature and some people who do not. Such data either explicitly provide, or can be manipulated to generate, a 2x2 classification of true positives and false negatives (in people who go on to develop the condition or outcome of interest) and false positives and true negatives (in people who do not).

When deciding whether data should be synthesised or presented separately, heterogeneity in the population, index feature and outcome to be predicted were considered to determine whether data could be meaningfully combined. When it was decided that data could be meaningfully combined, the same methods were used when synthesising predictive accuracy data as those described for synthesising diagnostic accuracy data.

Data synthesis for association data

In this guideline, association data were defined as measures of association between one or more factors (which could be either a single variable or a group of variables) and an outcome variable, where the data are not reported in terms of outcome classification (i.e. diagnostic/predictive accuracy). Examples could include (but were not limited to) data assessing the association between variables and diagnosis (diagnostic association studies) or data assessing the association between variables and a future outcome (prognostic association studies). Data were reported as hazard ratios (if measured over time) or odds ratios or risk ratios (if measured at a specific time-point). The committee agreed that odds ratios were the most appropriate method for understanding association data in this case.

The same methods for meta-analysis of odds ratios and relative risks were used as described as in the section on Data synthesis for intervention studies.

Data synthesis for qualitative reviews

Where multiple qualitative studies were identified for a single question, information from the studies was combined using a thematic synthesis. The thematic synthesis was based partly on a priori categories describing phenomena the committee was interested in (for example, using an existing model [framework synthesis]) and partly on themes that emerged from the coding of the included studies. Papers were uploaded to NVivo 11 software where the relevant data from the papers were coded. Once all of the included studies had been examined and coded, the resulting sets of codes were aggregated into themes and sub-themes. The aggregated themes were used to develop interpretive ‘review findings’ that were evaluated using CERQual. These review findings were reproduced in a summary of qualitative findings table along with example quotes and details of the CERQual assessment of each review finding.

Data synthesis for mixed methods reviews

Data synthesis for mixed methods reviews was carried out in accordance with the Joanna Briggs Institute manual for evidence synthesis (https://wiki.jbi.global/display/MANUAL) chapter 8. Synthesis followed a convergent segregated approach where independent synthesis of quantitative data and

qualitative data was undertaken, followed by the integration of the two types of evidence.

The qualitative and quantitative reviews were presented separately in the reviews and an integration section was written that addressed the following questions:

- Are the results/findings from individual syntheses supportive or contradictory?
- Does the qualitative evidence explain why the intervention is/is not effective?
- Does the qualitative evidence explain differences in the direction and size of effect across the included quantitative studies?
- Which aspects of the quantitative evidence were/were not explored in the qualitative studies?
- Which aspects of the qualitative evidence were/were not tested in the quantitative studies?

Where appropriate, and data from quantitative and qualitative sections of the review were integrated into tables or logic models/conceptual frameworks to show possible interrelationships between them.

Appraising the quality of evidence

Intervention studies (relative effect estimates)

RCTs and quasi-randomised controlled trials were quality assessed using the Cochrane Risk of Bias Tool. Non-randomised controlled trials and cohort studies were quality assessed using the ROBINS-I tool. Other study types (for example controlled before and after studies) were assessed using the preferred option specified in the NICE guidelines manual 2020 (appendix H). Evidence on each outcome for each individual study was classified into one of the following groups:

- **Low risk of bias** – The true effect size for the study is likely to be close to the estimated effect size.
- **Moderate risk of bias** – There is a possibility the true effect size for the study is substantially different to the estimated effect size.
- **High risk of bias** – It is likely the true effect size for the study is substantially different to the estimated effect size.
- **Critical risk of bias** (ROBINS-I only) - It is very likely the true effect size for the study is substantially different to the estimated effect size.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, intervention, comparator and/or outcomes in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- **Direct** – No important deviations from the protocol in population, intervention, comparator and/or outcomes.
- **Partially indirect** – Important deviations from the protocol in one of the following areas: population, intervention, comparator and/or outcomes.
- **Indirect** – Important deviations from the protocol in at least two of the following areas: population, intervention, comparator and/or outcomes.

Minimally important differences (MIDs) and decision thresholds

The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to identify published minimal important difference thresholds relevant to this guideline that might aid the committee in identifying decision thresholds for the purpose of GRADE. Identified MIDs were assessed to ensure they had been developed and validated in a methodologically rigorous way, and were applicable to the populations, interventions and outcomes specified in this guideline. In addition, PHAC members were asked to prospectively specify any outcomes where they felt a consensus decision threshold could be defined from their experience.

Decision thresholds were used to assess imprecision using GRADE.

For continuous outcomes expressed as a mean difference where no other decision threshold was available, a decision threshold of 0.5 of the median standard deviations of the comparison group arms was used (Norman et al. 2003). For continuous outcomes expressed as a standardised mean difference where no other decision threshold was available, a decision threshold of 0.5 standard deviations was used. For SMDs that were back converted to one of the original scales to aid interpretation, rating of imprecision was carried out before back calculation. For relative risks and hazard ratios, where no other decision threshold was available, a default decision threshold for dichotomous outcomes of 0.8 to 1.25 was used.

GRADE for pairwise meta-analyses of interventional evidence

GRADE was used to assess the quality of evidence for the outcomes specified in the review protocol. Data from randomised controlled trials, non-randomised controlled trials and cohort studies (which were quality assessed using the Cochrane risk of bias tool or ROBINS-I) were initially rated as high quality while data from other study types were initially rated as low quality. The quality of the evidence for each outcome was downgraded or not from this initial point, based on the criteria given in Table 2.

Table 2: Rationale for downgrading quality of evidence for intervention studies

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Extremely serious: If greater than 33.3% of the weight in a meta-analysis came from studies at critical risk of bias, the outcome was downgraded three levels
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels.

GRADE criteria	Reasons for downgrading quality
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 50%, the outcome was not downgraded. Serious: If the I^2 was between 50% and 75%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 75%, the outcome was downgraded two levels.
Imprecision	The line of no effect was considered to be a key decision point for imprecision for all outcomes in these reviews. An outcome was downgraded once if the 95% confidence interval for the effect size crossed the line of no effect (i.e. the outcome was not statistically significant). Outcomes were downgraded twice if the confidence intervals crossed both MIDs as described above.
Publication bias	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. When a funnel plot showed convincing evidence of publication bias, or the review team became aware of other evidence of publication bias (for example, evidence of unpublished trials where there was evidence that the effect estimate differed in published and unpublished data), the outcome was downgraded once. If no evidence of publication bias was found for any outcomes in a review (as was often the case), this domain was excluded from GRADE profiles to improve readability.

For outcomes that were originally assigned a quality rating of 'low' (when the data was from observational studies that were not appraised using the ROBINS-I checklist), the quality of evidence for each outcome was upgraded if any of the following three conditions were met and the risk of bias for the outcome was rated as 'no serious':

- Data from studies showed an effect size sufficiently large that it could not be explained by confounding alone.
- Data showed a dose-response gradient.
- Data where all plausible residual confounding was likely to increase our confidence in the effect estimate.

Association studies

Individual prognostic studies presenting data on association were quality assessed using the QUIPs checklist. Other cohort and case-control studies were quality assessed using the CASP cohort study and case-control checklists, respectively. Individual cross-sectional studies were quality assessed using the Joanna Briggs Institute critical appraisal checklist for analytical cross-sectional studies (2016), which contains 8 questions covering: inclusion criteria, description of the sample, measures of exposure, measures of outcomes, confounding factors, and statistical analysis. Each study was classified into one of the following groups:

- **Low risk of bias** – The true effect size for the study is likely to be close to the estimated effect size.

- **Moderate risk of bias** – There is a possibility the true effect size for the study is substantially different to the estimated effect size.
- **High risk of bias** – It is likely the true effect size for the study is substantially different to the estimated effect size.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, factors and/or outcomes in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- **Direct** – No important deviations from the protocol in population, factors and/or outcomes.
- **Partially indirect** – Important deviations from the protocol in one of the population, factors and/or outcomes.
- **Indirect** – Important deviations from the protocol in at least two of the population, factors and/or outcomes.

Public health decision thresholds

The committee were asked to define decision thresholds for association outcomes based on the degree of association that was considered important for decision making. In cases where the committee were unable to define a decision threshold by consensus, the line of no effect was used at the decision threshold for the purpose of rating imprecision in GRADE.

Modified GRADE for association data

GRADE has not been developed for use with association studies. The data from the association studies included in these reviews mostly reported adjusted odds ratios for outcomes but with no raw data it was not possible to meta-analyse the data or to apply the modified approach to using the GRADE framework that is recommended in the NICE manual.

GRADE-CERQual for qualitative evidence synthesis findings

CERQual was used to assess the confidence we have in each of the review findings. Evidence from all qualitative study designs (interviews, focus groups etc.) was initially rated as high confidence and the confidence in the evidence for each theme was assessed from this initial point as detailed in Table 9 below. Confidence in each criterion was assessed as:

- No or very minor concerns
- Minor concerns
- Moderate concerns
- Serious concerns

And an overall confidence rating of High, Moderate, Low or Very Low was determined based on this.

Table 3 Overall confidence in qualitative outcome

Level	Definition
High confidence	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest

Level	Definition
Moderate confidence	It is likely that the review finding is a reasonable representation of the phenomenon of interest
Low confidence	It is possible that the review finding is a reasonable representation of the phenomenon of interest
Very low confidence	It is not clear whether the review finding is a reasonable representation of the phenomenon of interest

Table 9 Rationale for downgrading confidence in evidence for qualitative questions

CERQual criteria	Reasons for downgrading confidence
Methodological limitations	One or more studies contribute data to each review finding in a qualitative evidence synthesis, and these data make up the body of data for a review finding. The methodological limitations of the body of data supporting a review finding are assessed as a whole to identify whether or not any methodological weaknesses within individual studies impact our confidence in a review finding. The methodological limitations for each review finding must be assessed separately since different studies contribute varying amounts of data to each review finding, and methodological quality issues may have varying impacts on different review findings.
Relevance	Relevance is the extent to which the body of data from the primary studies supporting a review finding is applicable to the context specified in the review question. Relevance is the CERQual component that is anchored to the context specified in the review question. How the review question and objectives are expressed, how a priori subgroup analyses are specified and how theoretical considerations inform the review design are therefore critical to making an assessment of relevance when applying CERQual.
Coherence	The coherence of a review finding is an assessment of how clear and cogent the fit is between the data from the primary studies and a review finding that synthesises that data. It includes consideration of the general 'fit' of data and whether any discrepancies can be explained.
Adequacy of data	Adequacy of data is an overall determination of the degree of richness as well as the quantity of data supporting a review finding. • Richness of the data is the extent to which the information that the individual study authors have provided is detailed enough to allow the review author to interpret the meaning and context of what is being researched. • Quantity of data relates to the number of studies and participants that this data comes from.

Mixed methods studies

Mixed methods studies were evaluated using the appropriate quality assessment tools for the component study types, see sections on intervention studies and qualitative studies. Other methods of assessing mixed methods studies were agreed

with the NICE methods and economics team QA lead and reported in the individual reviews.

Reviewing economic evidence

Inclusion and exclusion of economic studies

Literature reviews seeking to identify published cost effectiveness studies of relevance to the issues under consideration were conducted for all questions that had an effectiveness component. In each case, the search undertaken for the public health review was modified, retaining population and intervention descriptors, but removing any study-design filter and adding a filter designed to identify relevant health economic analyses. In assessing studies for inclusion, population, intervention and comparator, criteria were always identical to those used in the parallel public health search; only comparative cost effectiveness analyses were included. Economic evidence profiles, including critical appraisal according to the Guidelines manual, were completed for included studies.

Appraising the quality of economic evidence

Economic studies identified through a systematic search of the literature were appraised using a methodology checklist designed for economic evaluations (NICE guidelines manual; 2020). It is used to determine whether an economic evaluation provides evidence that is useful to inform the decision-making of the Committee. It judges the applicability of the study and the limitations.

There are 2 parts of the appraisal process. The first step is to assess applicability (that is, the relevance of the study to the specific guideline topic and the NICE reference case); evaluations are categorised according to the criteria in Table 10.

Table 10 Applicability criteria

Level	Explanation
Directly applicable	The study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness
Partially applicable	The study fails to meet one or more applicability criteria, and this could change the conclusions about cost effectiveness
Not applicable	The study fails to meet one or more applicability criteria, and this is likely to change the conclusions about cost effectiveness. These studies are excluded from further consideration

In the second step, only those studies deemed directly or partially applicable are further assessed for limitations (that is, methodological quality); see categorisation criteria in Table .

Table 11 Methodological criteria

Level	Explanation
Minor limitations	Meets all quality criteria, or fails to meet one or more quality criteria but this is unlikely to change the conclusions about cost effectiveness
Potentially serious limitations	Fails to meet one or more quality criteria and this could change the conclusions about cost effectiveness

Level	Explanation
Very serious limitations	Fails to meet one or more quality criteria and this is highly likely to change the conclusions about cost effectiveness. Such studies should usually be excluded from further consideration

Where relevant, a summary of the main findings from the systematic search, review and appraisal of economic evidence is presented in an economic evidence profile alongside the public health evidence.

Health economic modelling

As well as reviewing the published economic literature for each effectiveness review question, as described above, de novo economic analysis was undertaken in selected areas. Priority areas for new health economic analysis were agreed by the committee.

The following general principles were adhered to in developing the analysis:

- Methods were consistent with the NICE reference case.
- The design of the model, selection of inputs and interpretation of the results was discussed and agreed with the committee.
    - Where possible, model inputs were based on the systematic review of the public health literature, supplemented with other published data sources identified by the committee as required.
    - When published data were not available committee expert opinion was used to populate the model.
    - Model inputs and assumptions were reported fully and transparently.
    - The results were subject to sensitivity analysis and limitations were discussed.

Full methods for the de novo cost-effectiveness analysis are described in the HE report.

Resource impact assessment

The resource impact team used the methods outlined in the in Assessing resource impact process manual: guidelines

The resource impact team worked with the guideline committee from an early stage to identify recommendations that either individually or cumulatively would a substantial impact on resources. The aim was to ensure that a recommendation would not introduce a cost pressure into the health and social care system unless the committee was convinced of the benefits and cost effectiveness of the recommendation. The team gave advice to the committee on issues related to the workforce, capacity and demand, training, facilities and educational implications of the recommendations.