National Institute for Health and Care Excellence

DRAFT FOR CONSULTATION

Tobacco: preventing uptake, promoting quitting and treating dependence (update)

NICE guideline NGXX

Methods

June 2021

NICE guideline: methods

DRAFT FOR CONSULTATION

Evidence reviews were developed by the Public Health Internal Guideline Development 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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: XX

Contents

Development of the guideline

What this guideline covers

This guideline covers the methods used for the updated and new sections of Tobacco: preventing uptake, promoting quitting and treating dependence (update).

This guideline brings together NICE guidelines PH5, PH14, PH23, PH26, PH39, PH45, PH48 and NG92 and will replace them. We have reviewed evidence on:

- digital mass media for preventing uptake
- cessation mass media for preventing uptake
- proxy purchasing and supply of illicit tobacco
- impact of e-cigarettes on future smoking behaviour
- smokefree class competitions for preventing uptake
- opt-out referral to stop smoking support in pregnancy
- incentives for cessation in pregnancy
- effectiveness, safety and acceptability of NRT and e-cigarettes for cessation in pregnancy
- effectiveness of treatments for cessation
- barriers and facilitators to using e-cigarettes for cessation
- long-term health effects of using e-cigarettes
- relapse prevention
- tailored interventions for those with mental health conditions

Recommendations are marked [2021] if the evidence has been reviewed.

What this guideline does not cover

This guideline does not cover the methods used for sections of the guideline which are out of scope or carried forward from previous guidelines. Those sections are labelled with the publication date of the previous guidelines.

1 Methods

This guideline was developed in accordance with the process set out in 'Developing NICE guidelines: the manual (2018)'. Where the guidelines manual does not provide advice, additional methods are described below.

1.1 Developing the review questions and outcomes

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 following framework:

- Population, intervention, comparator and outcome (PICO) for reviews of interventions
- Sample, phenomena of interest, design, evaluation, research type (SPIDER) for qualitative reviews

For all review questions, the following were completed as appropriate: full literature searches, lists of excluded studies and reasons for exclusion, evidence tables and critical appraisal for all included studies.

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 review.

1.2 Priority screening

The reviews undertaken for this guideline all made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software (EPPI-4). This 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, where search results were large (over 5,000 results):

- At least 50% of the total identified abstracts were screened AND
- After this point, if at least 10% of the total identified abstracts were sifted without identifying a potential include

To ensure this approach did not miss relevant studies, the included study lists of included systematic reviews were searched to identify any papers not identified through the primary search. The study inclusion and exclusion lists were viewed by the PHAC on completion of the review to ensure no studies are excluded inappropriately.

1.3 Incorporating published systematic reviews

1.3.1 Using Cochrane reviews

Cochrane reviews which were identified as directly relevant to a review were used in different ways depending on when they were identified. If identified at an early stage, for example during scoping or protocol drafting, Cochrane were approached to update their review and provide NICE with the results. If identified at a later stage, the published Cochrane results were used in combination with data extracted from more recent publications identified through NICE's systematic searches. Results were combined and synthesised together.

Risk of bias

The ROBIS risk of bias tool was used to assess the Cochrane review. The review was classified into one of the following three groups:

- High quality – It is unlikely that additional relevant and important data would be identified from primary studies compared to that reported in the review, and unlikely that any relevant and important studies have been missed by the review.
- Moderate quality – It is possible that additional relevant and important data would be identified from primary studies compared to that reported in the review, but unlikely that any relevant and important studies have been missed by the review.
- Low quality – It is possible that relevant and important studies have been missed by the review.

Inconsistency

There are various methods of dealing with inconsistency in meta-analysis. Where the methods used in a Cochrane review differed from the method agreed for this guideline, amendments were made to bring in line with this guideline. See section 1.4.4.

GRADE

GRADE tables will be produced using data extracted from the Cochrane reviews and applying rules about the individual domains (for example, MIDs, inconsistency thresholds).

1.4 Evidence of effectiveness of interventions

1.4.1 Obtaining effect estimates for continuous outcomes

Where possible, meta-analyses were conducted to combine the results of quantitative studies for each outcome. For continuous outcomes analysed as mean differences, where change from baseline data were reported in the trials and were accompanied by a measure of spread (for example standard deviation), these were extracted and used in the meta-analysis.

1.4.2 Obtaining effect estimates for binary outcomes

Risk ratios (RR) were the preferred relative effect estimate for this guideline. Where studies present the odds ratio, this was converted to a risk ratio. The event rate in the control arm was used as the prevalence in the calculation. Study-level risk ratios

were then combined in meta-analysis. For the Network Meta-Analysis (NMA, see section 1.8) event data was modelled as odds ratios and the pooled odds ratio was converted to a risk ratio.

Where confidence intervals are not reported for effect estimates, the P-Value and point estimate are used to derive the confidence intervals using RevMan[1].

1.4.3 Methods for combining intervention evidence

Meta-analyses of interventional data were conducted with reference to the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2011). Meta-analyses were performed in Cochrane Review Manager v5.3.

Dichotomous outcomes from cluster and individual randomised controlled trials were pooled (using the Mantel–Haenszel method).

Randomised and non-randomised trials investigating the same outcomes were not pooled, in line with Cochrane recommendations not to pool these two study types.

For binary data, absolute risks were also presented in GRADE. Absolute risks were calculated based on the baseline risk (or 'control group risk': number with the event in the control group divided by total number in the control group), and the relative effect size. Where multiple studies are combined, control groups were summed and averaged using GRADEpro and expressed per 1000.

Where non-randomised studies had already conducted adjustments for confounding or for clustering (in studies where allocation was by cluster), the adjusted effect estimate was retained.

1.4.4 Heterogeneity

Where significant between study heterogeneity in methodology, population, intervention or comparator was identified in advance of data analysis but pooling was deemed appropriate, random effects models were used. If the committee decided that it was inappropriate to pool studies due to excess contextual or methodological differences, studies were not pooled.

In all other cases, a fixed effects model was preferred. If a fixed effect model was initially applied but the I^2 value was ≥50%, it was decided that the assumption of a shared mean was clearly not met and so a random effects models was then applied .

Where subgroups had been prespecified in the protocol, the I^2 was used as a test for subgroup differences. Where <50%, it was considered that there were no subgroup differences, and so the overall result was used. Where subgroup differences were present (I^2 was ≥50%), the subgroups were presented separately and the appropriate fixed or random effects models used dependent on the I^2 for the studies within that subgroup.

In any meta-analyses where there was significant heterogeneity (≥50% I^2) not resolved by subgroups and some (but not all) of the data came from studies at high risk of bias, a sensitivity analysis was conducted, excluding those studies from the analysis. Results from both of these meta-analyses are reported.

[1] Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

1.4.5 Clustering

For studies where allocation was by cluster and not individual, studies should correct for the effect of clustering. If no adjustment had been carried out, the review team adjusted the effect estimates by inflating standard errors, as described in the Cochrane manual (Higgins et al. 2011).

Intracluster correlation coefficients (ICCs), required to inflate the standard errors, were initially sought from a similar included study. Where there were no ICCs available in included studies, the literature was searched more broadly for previously used examples. For example, an ICC of 0.075 was used in reviews about classroom-based prevention programmes, identified from an included study[2].

1.4.6 Minimal clinically important differences (MIDs)

MIDs were identified through consultation with the committee, based on their expertise and what they would consider to be a meaningful change.

No MID thresholds relevant to this guideline were identified from the COMET database or other published source. Where not decided otherwise, default MIDs were applied. Default MID thresholds for dichotomous outcomes are a relative risk of 0.8 and 1.25. For continuous outcomes, default MIDs are 0.5*standard deviation (SD) of the control group.

If a continuous MID could not be calculated (e.g. where the committee did not agree an MID and so the default was used, and the standard deviation of the outcome measure at baseline was not reported in the paper) then we downgraded by 1 level as it was 'not possible to calculate imprecision from the information reported in the study'. Where data was pooled in analyses, the study with the largest weight was used as the control group for default MID calculations.

The committee considered the clinical importance in their discussion of the reviews presented.

1.4.7 GRADE for pairwise meta-analyses of interventional evidence

GRADE was used to assess the certainty of evidence for the selected outcomes as specified in 'Developing NICE guidelines: the manual (2014)'.

Effectiveness reviews

Data from all RCTs were initially rated as high quality and downgraded as necessary (see Table 1). Data from all studies which were assessed using the ROBINS I checklist were initially rated as high quality and downgraded as necessary. Data from studies which were not randomised or rated using the ROBINS I checklist were initially rated as low quality to take into account the risks of bias inherent to non-randomised studies (and not already accounted for in the risk of bias checklist). These were only downgraded for risk of bias in GRADE where there were clear and significant additional risks.

[2] M R Crone, S A Reijneveld, M C Willemsen et al., 2003. Prevention of smoking in adolescents with lower education: a school based intervention study. *Journal of Epidemiology and Community Health,* 57:675-680.

Other review types

Reviews for which non-randomised studies may be one of the best sources of evidence, for example prognostic reviews, will use GRADE starting points as follows: all appropriate study designs will start as high and will be downgraded via the GRADE domains, including risk of bias.

Table 1: Rationale for downgrading quality of evidence for intervention studies

GRADE domain	Description
Risk of bias	Limitations in study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the study subject, healthcare professional or assessor) and attrition bias (due to missing data causing systematic bias in the analysis). Sometimes randomisation may not adequately lead to group equivalence of confounders, and if so, this may lead to bias, which should be taken into account. Where there are no study limitations, evidence is assessed as having 'no serious' risk of bias. Alternatively, evidence may be downgraded one level ('serious' risk of bias) or two levels ('very serious' risk of bias).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question, which might affect the effect estimate. Where the evidence is directly applicable to the PICO, it is assessed as having 'no serious' risk of indirectness. Alternatively, evidence may be downgraded one level ('serious' risk of indirectness) or two levels ('very serious' risk of indirectness).
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies pooled in the same meta-analysis. The I^2 statistic describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance). Where the I^2 value is above 50% heterogeneity will be judged to be serious and so will be downgraded by one level in GRADE. It will also be explored by using pre-specified subgroup analysis. If the I^2 value is above 75%, heterogeneity will be judged to be very serious.
Imprecision	Results are imprecise when studies include relatively few participants and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example a result may be consistent with both public health benefit and public health harm) and thus be imprecise. Uncertainty is introduced where confidence intervals cross the MID threshold. If the confidence interval crosses one MID threshold, this indicates 'serious' risk of imprecision in the effect estimate. Crossing both MID thresholds indicates 'very serious' risk of imprecision in the effect estimate. Where the MID is 'any significant change' there is effectively only one threshold (the line of no effect), and so only one opportunity for downgrading. In this instance, outcomes will be

GRADE domain	Description
	downgraded again if they are based on small samples (<300 participants across both intervention and comparator groups).
Other issues	Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely related phenomenon is where some papers fail to report an outcome that is inconclusive, thus leading to an overestimate of the effectiveness of that outcome.

Table 2: Overall GRADE rating meanings

Overall GRADE rating	Description
High	Further research is very unlikely to change our confidence in the estimate of effect.
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very Low	Any estimate of effect is very uncertain.

1.4.82 Publication bias

Publication bias was assessed 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.

1.4.96 Evidence summaries

GRADE profiles provide full details of results. Evidence summaries are intended to replace evidence statements, and to provide a high-level overview to summarise GRADE profiles.

Evidence summaries contain the following information in table format:

- Outcome name
- Summary statement (including relevant studies)
- Certainty (high, moderate, low, very low)
- Related GRADE profile number

Summary statements were written as follows where the MID was the line of no effect

Summary statement	Meaning
The result suggests a clinically important increase / decrease in the outcome in X compared with Y.	Where the confidence interval does not include the MID
The result suggests no clinically important difference in the outcome in X compared with Y	Where the confidence interval includes the MID

Summary statements were written as follows where the MID was other than the line of no effect:

Summary statement	Meaning
The result suggests a clinically important increase / decrease in the outcome in X compared with Y	The point estimate exceeds the MID and the confidence interval does not include either MID
The result suggests a clinically important increase / decrease in the outcome in X compared with Y, with some / high uncertainty	The point estimate exceeds the MID, but the confidence interval includes one / both MIDs.
The result suggests no clinically important difference in the outcome in X compared with Y	The point estimate does not exceed the MID and the confidence interval does not include either MID.
The result suggests no clinically important difference in the outcome in X compared with Y, with some / high uncertainty	The point estimate does not exceed the MID, but the confidence interval includes one / both MIDs.

Additional detail was added as needed about subgroup or sensitivity analysis.

1.52 Qualitative evidence

Where multiple qualitative studies were identified for a single question, information from the studies was combined using a thematic synthesis. The aggregated themes were used to develop interpretive 'review findings'. These review findings were assessed using GRADE CERQual and are presented in review documents.

GRADE CERQual

CERQual was used to assess the confidence in each of the review findings from qualitative studies. Evidence from all qualitative study designs (interviews, focus groups etc.) was initially rated as high confidence and the confidence in the evidence for each finding was then downgraded from this initial point based on quality assessment of the studies.

Representative quotations are presented alongside the review findings.

For each domain, concerns were rated as: no or very minor; minor; moderate; or serious. Table 3 details how these judgements were made.

Table 3: CERQual

Criterion	Reason for downgrading or not downgrading
Methodological limitations	No or very minor: risks of bias are none / minimal Minor: risks to bias are unlikely to have an effect on the finding Moderate: risks to bias may change the finding Serious: risks to bias may significantly change the finding Methodological limitations are informed by the risk of bias of the study, and whether the risk of bias is likely to affect the finding under consideration.
Coherence	No or very minor: the fit between the data and the review finding is very good Minor: there are some minor concerns that the fit between the data and the review finding may be flawed Moderate: There are significant discrepancies between the finding and the data Serious: There are very significant discrepancies between the finding and the data

Criterion	Reason for downgrading or not downgrading
	This is informed by whether the finding accurately reflects all relevant data from all relevant studies.
Adequacy of data	No or very minor: Data is rich and comes from multiple studies Minor: Data is fairly rich and comes from multiple studies Moderate: Data is not rich or comes from few studies Serious: Data is not rich and comes from few studies The outcome was downgraded if there was insufficient data to develop an understanding of the phenomenon of interest, either due to insufficient studies or observations.
Relevance	No or very minor: finding is from highly relevant studies Minor: finding is from relevant and partially relevant studies Moderate: finding is from partly relevant studies only Serious: finding is from studies with serious issues to do with relevance, or from a variety of studies. Data may have relevance issues where it is indirectly relevant, where it is only partially relevant, or where its relevance is unclear.

Table 4: Overall CERQual rating meanings

Overall GRADE rating	Description
High	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest.
Moderate	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 not clear whether the review finding is a reasonable representation of the phenomenon of interest.

1.6 Combining qualitative and quantitative evidence

A matrix method was used to juxtapose qualitative with quantitative findings (Harden, 2018) in order to integrate the two types of data. Qualitative findings were viewed alongside quantitative findings to explore whether they explained the effects seen. These comparisons were used to aid committee discussion.

1.7 Quality appraisal

Quality assessment for all included outcomes was conducted using the tools in Developing NICE guidelines: the manual. Checklists were chosen according to review type and study design.

Table 5: Checklists

Review type	Study design	Critical appraisal checklist
Effectiveness	Systematic Review	ROBIS
	RCT	Cochrane RoB tool (2.0)
	cRCT	Cochrane RoB tool (2.0) for cluster trials

Review type	Study design	Critical appraisal checklist
	NRS (incl. cohort)	Cochrane ROBINS-I
Association	Systematic Review	ROBIS
	Cohort or ITS	QUIPS
Qualitative	Systematic Review	ROBIS
	Any other	CASP qualitative checklist

RCT: Randomised controlled trial
cRCT: cluster RCT
NRS: non-randomised study

NICE's statement on engagement with tobacco industry organisations was followed for this guideline. Potential conflicts of interest were carefully considered by the committee.

1.8 Methods for combining direct and indirect evidence (network meta-analysis) for interventions

In situations where there are more than two interventions, pairwise meta-analysis of the direct evidence alone is of limited use. This is because multiple pairwise comparisons need to be performed to analyse each pair of interventions in the evidence, and these results can be difficult to interpret. Furthermore, direct evidence about interventions of interest may not be available. For example, studies may compare A vs B and B vs C, but there may be no direct evidence comparing A vs C. Network meta-analysis overcomes these problems by combining all evidence into a single, internally consistent model, synthesising data from direct and indirect comparisons, and providing estimates of relative effectiveness for all comparators and the ranking of different interventions. Network meta-analyses were undertaken where the following three criteria were met:

- At least three treatment alternatives.
- A sufficiently connected network to enable valid estimates to be made.

1.8.1 Synthesis

WinBUGS version 1.4.3 was used to perform Hierarchical Bayesian Network Meta-Analysis (NMA) for this guideline. The models used reflected the recommendations of the NICE Decision Support Unit's Technical Support Documents (TSDs) on evidence synthesis, particularly TSD 2 ('A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials'; see http://www.nicedsu.org.uk). As an existing NMA was being conducted by a team from Bristol University in an area under investigation for this guideline, the model used for that NMA was used and rerun after data amended for the purposes of the review.

Specifics of methodology used for the NMA included in this guideline are outlined in Appendix I of review K: cessation and harm reduction treatments.

Fixed- and random-effects models were explored for cessation at 6 months, with the final choice of model based on deviance information criterion (DIC): if DIC was at least 3 points lower for the random-effects model, it was preferred; otherwise, the fixed effects model was considered to provide an equivalent fit to the data in a more parsimonious analysis and was preferred.

In any network meta-analyses where some (but not all) of the data came from studies at high risk of bias, a sensitivity analysis was conducted, excluding those studies

from the analysis. Results from both the full and restricted network meta-analyses are reported. If an external review was being used or updated and had already conducted this analysis finding no significant difference, this was not conducted.

1.8.24 GRADE for network meta-analyses

GRADE was applied to the pairwise meta-analysis data inputted into the NMA in the same way as for all other pairwise data in this guideline (see section 1.4.7).

A modified version of the standard GRADE approach for pairwise interventions was used to assess the quality of evidence across the network meta-analyses undertaken. While most criteria for pairwise meta-analyses still apply, it is important to adapt some of the criteria to take into consideration additional factors, such as how each 'link' or pairwise comparison within the network applies to the others. As a result, the following was used when modifying the GRADE framework to a network meta-analysis. It is designed to provide a single overall quality rating for an NMA, which can then be combined with pairwise quality ratings for individual comparisons (if appropriate), to judge the overall strength of evidence for each comparison.

Table 6: Rationale for downgrading quality of evidence for intervention studies

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If fewer than 33.3% of the studies in the network meta-analysis had some concerns or high risk of bias, the overall network was not downgraded. Serious: If greater than 33.3% of the studies in the network meta-analysis had some concerns or high risk of bias, the network was downgraded one level. Very serious: If greater than 33.3% of the studies in the network meta-analysis were at high risk of bias, the network was downgraded two levels.
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question which might affect the effect estimate. Where the evidence is directly applicable to the PICO, it is assessed as having 'no serious' risk of indirectness. Alternatively, evidence may be downgraded one level ('serious' risk of indirectness) or two levels ('very serious' risk of indirectness).
Inconsistency	N/A: Inconsistency was marked as not applicable if there were no links in the network where data from multiple studies (either direct or indirect) were synthesised. For network meta-analyses conducted under a Bayesian framework, the network was downgraded one level if the DIC for a random-effects model was lower than the DIC for a fixed-effects model. In addition the direct and indirect treatment estimates were compared as a check on the consistency of the network.
Imprecision	Not serious: The data were sufficiently precise by the committee to meet the aims of the review question. Serious: Imprecision had a moderate impact on the ability of the data to meet the aims of the review question. Very serious: Imprecision had a substantial impact on the ability of the data to meet the aims of the review question.

1.98 Association reviews

In this guideline, association reviews are present to determine the association between a factor and an outcome. Study outcomes were extracted to provide

information on the adjusted relative risks of developing the outcome in those with the predictive factor compared with those without the predictive factor.

Rationale for downgrading the quality of evidence for association questions was as for intervention evidence, with the exception that adjustments for pre-specified confounders will be considered to be especially necessary to reduce risk of bias in the outcome.

Association studies considered outcomes which the committee decided should have MIDs equal to the line of no effect.

Table 7: Evidence summaries for association reviews

Summary statement	Meaning
The result suggests no association between exposure and outcome.	The CI includes the line of no effect.
The result suggests an association between exposure and a clinically important increase / decrease in the outcome.	The CI does not include the line of no effect.

Additional detail will be added as needed about subgroup or sensitivity analysis.

1.10 Health economics

Literature reviews seeking to identify published cost–utility analyses of relevance to the issues under consideration were conducted for all intervention effectiveness questions. In each case, the search undertaken for the effectiveness 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.

York Health Economics Consortium conducted cost effectiveness reviews. Details of methodology are presented within the evidence reviews.

Reviewing economic evidence

The public health advisory committee is required to make decisions based on the best available evidence of both effectiveness and cost-effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost-effectiveness') rather than the total implementation cost. Thus, if the evidence suggests that a strategy provides significant health benefits at an acceptable cost per recipient, it should be recommended.

In order to assess the cost effectiveness of the key issues addressed in this guideline, the following actions were carried out:

- A systematic review of economic evidence in the literature was conducted, alongside the review of evidence on effectiveness
- A *de novo* economic model was developed, in order to provide cost effectiveness evidence for a number of review questions prioritised by the committee

Inclusion and exclusion of economic studies

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost-utility, cost-effectiveness, cost-benefit and cost-consequence analyses) and comparative costing studies that addressed the review question in the relevant population were considered potentially includable as economic evidence.

As per 'Developing NICE Guidelines: The Manual', UK-based cost-utility studies reporting health outcomes in quality adjusted life years (QALYs) were preferred. However, due to the relatively sparse evidence for most review questions, non-UK-based cost effectiveness studies (i.e. those reporting outcomes in natural units, such as number of successful quitters) were also included. It was determined that such evidence may still be useful in informing the committee of the potential trade-off between costs and benefit of interventions. Similarly, cost-consequence analyses (i.e. those in which costs and benefits are reported separately) were included, as they include both health and non-health effects and tend to take a broad perspective making them particularly relevant to decision makers in different settings, such as local authorities.

Studies which only reported costs (without any consideration of health benefits) were excluded. Literature reviews, abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded.

Appraising the quality of economic evidence

Studies that met the eligibility criteria were assessed using the quality appraisal checklist for economic evaluations set out in Appendix H of Developing NICE guidelines: the manual (NICE 2018).

Health economic modelling

As well as reviewing the published economic literature for each 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 clinical 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.