National Institute for Health and Care Excellence

Final

Vaccine uptake in the general population

[L] NICE guideline: methods

NICE guideline NG218

Methods

May 2022

Final

This evidence review was developed by the 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-4587-0

Contents

Development of the guideline **5**
Remit 5
What this guideline covers 5
Vaccines of interest 5
Groups of interest 5
Settings of interest 5
Activities, services or aspects of care 5
What this guideline does not cover 5
Methods **7**
Developing the review questions and outcomes 7
Call for evidence 7
Reviewing research evidence 8
Type of studies and inclusion/exclusion criteria 8
Methods of combining evidence 10
Data synthesis for intervention studies 10
Data synthesis for qualitative studies 11
Data synthesis for mixed methods studies and reviews 12
Appraising the quality of evidence 13
Intervention studies 13
Qualitative studies 15
Mixed methods studies 16
Reviewing economic evidence 17
Inclusion and exclusion of economic studies 17
Appraising the quality of economic evidence 17
Costing exercises 18
References 19

Development of the guideline

Remit

The Department of Health and Social Care in England has asked NICE to produce a guideline on vaccine uptake in the general population.

This guideline will update and replace the NICE guideline on immunisations: reducing differences in uptake in under 19s (PH21).

This guideline will also be used to develop the NICE quality standard for vaccine uptake in the general population.

What this guideline covers

Vaccines of interest

Routine vaccines refers to vaccines available on the UK immunisation schedule as mentioned in chapter 11: the UK immunisation schedule of the Green Book. For this guideline it excludes seasonal flu vaccine.

Groups of interest

All people who are eligible for vaccines on the routine UK immunisation schedule. Specific consideration will be given to the groups listed in the equality impact assessment.

Settings of interest

- All settings where routine UK immunisation schedule vaccines are offered or delivered.
- Occupational health services.
- Education settings, including early years settings, schools, pupil referral units and universities.
- Private health clinics and vaccination centres where NHS-funded care is delivered.
- Secure settings, including prisons and immigration removal centres.

Activities, services or aspects of care

We will look at evidence in the areas below when developing the guideline, but it may not be possible to make recommendations in all the areas.

1. Identifying and recording a person's vaccination eligibility and status.
2. Increasing the uptake of routine vaccines.

What this guideline does not cover

- Areas covered by NICE's guideline on tuberculosis.
- Areas covered by NICE's guideline on flu vaccination: increasing uptake.

- Travel vaccines.
- Selective immunisation programmes, as defined in the Green Book.
- Seasonal vaccinations, for example flu vaccination.
- Catch-up campaigns alongside the introduction of a new vaccine.

Methods

This guideline was developed using the methods described in the 2018 NICE guidelines manual.

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

Developing the review questions and outcomes

The 11 review questions developed for this guideline were based on the key areas identified in the guideline scope. They were drafted by the Guideline Updates 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
- Sample, Phenomenon of Interest, Design, Evaluation, Research type (SPIDER) for qualitative reviews.

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

Call for evidence

A call for evidence was used to try to identify evidence that answered the following question:

- Is there any learning from the COVID-19 vaccination program that could be used to increase uptake of routine vaccines?

We requested the following information:

- Qualitative evidence focusing the barriers and facilitators to COVID- 19 vaccination uptake.
- Quantitative evidence about the effectiveness of interventions aimed at increasing COVID-19 vaccine uptake.

This approach was taken rather than carrying out a systematic review of evidence for the 2 areas above because we expected that there would be limited published evidence available at the time the call for evidence was made due to how recently the COVID-19 vaccination programme had been started.

We requested published or unpublished information meeting the above criteria, including any ongoing research. We did not accept promotional material, non-evidence-based assertions of effectiveness or opinion pieces.

The results of the call for evidence are presented in document K. This contains the methods used to select the studies of interest and how they were analysed.

Reviewing research evidence

Type of studies and inclusion/exclusion criteria

Searching for evidence

Evidence was searched for each review question using the methods specified in the 2018 NICE guidelines manual.

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.

The reviews undertaken for this guideline all made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. 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. Priority screening was used for each review, but no criteria was set for stopping abstract screening. Consequently, the whole abstract database was searched.

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.

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.

If published evidence syntheses were identified sufficiently early in the review process (for example, from the surveillance review or early in the database search), they were considered for use as the primary source of data, rather than extracting information from primary studies. Syntheses considered for inclusion in this way were quality assessed to assess their suitability using the appropriate checklist, as outlined in Table 1. Note that this quality assessment was solely used to assess the quality of the synthesis in order to decide whether it could be used as a source of data, as

outlined in Table 2, not the quality of evidence contained within it, which was assessed in the usual way as outlined in the section on 'Appraising the quality of evidence'.

Table 1 Checklist for published evidence syntheses

Type of synthesis	Checklist for quality appraisal
Systematic review of quantitative evidence	ROBIS
Qualitative evidence synthesis	ENTREQ reporting standard for published evidence synthesis is the generic reporting standard for QES, however specific reporting standards exist for meta-ethnography (eMERGe) and for realist synthesis (RAMESES II. If these reporting standards are not appropriate to the QES then an adapted PRISMA framework is used (see Flemming K, Booth A, Hannes K, Cargo M, Noyes J. Cochrane Qualitative and Implementation Methods Group guidance series-paper 6: reporting guidelines for qualitative, implementation, and process evaluation evidence syntheses. Journal of Clinical Epidemiology 2018; 97: 79-85).

Each published evidence synthesis 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.

Each published evidence synthesis was also classified into one of three groups for its applicability as a source of data, based on how closely the review matches the specified review protocol in the guideline. Studies were rated as follows:

- Fully applicable – The identified review fully covers the review protocol in the guideline.
- Partially applicable – The identified review fully covers a discrete subsection of the review protocol in the guideline (for example, some of the factors in the protocol only).
- Not applicable – The identified review, despite including studies relevant to the review question, does not fully cover any discrete subsection of the review protocol in the guideline.

The way that a published evidence synthesis was used in the evidence review depended on its quality and applicability, as defined in Table 2. When published evidence syntheses were used as a source of primary data, data from these evidence syntheses were quality assessed and presented in GRADE/CERQual tables in the same way as if data had been extracted from primary studies. In questions where data was extracted from both systematic reviews and primary studies, these were checked to ensure none of the data had been double counted through this process.

Table 2 Criteria for using published evidence syntheses as a source of data

Quality	Applicability	Use of published evidence synthesis
High	Fully applicable	Data from the published evidence synthesis were used instead of undertaking a new literature search or data analysis. Searches were only done to cover the period of time since the search date of the review. If the review was considered up to date (following discussion with the guideline committee and NICE lead for quality assurance), no additional search was conducted.
High	Partially applicable	Data from the published evidence synthesis were used instead of undertaking a new literature search and data analysis for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. If the review was considered up to date (following discussion with the guideline committee and NICE lead for quality assurance), no additional search was conducted. For other sections not covered by the evidence synthesis, searches were undertaken as normal.
Moderate	Fully applicable	Details of included studies were used instead of undertaking a new literature search. Full-text papers of included studies were still retrieved for the purposes of data analysis. Searches were only done to cover the period of time since the search date of the review.
Moderate	Partially applicable	Details of included studies were used instead of undertaking a new literature search for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. For other sections not covered by the evidence synthesis, searches were undertaken as normal.

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. Network meta-analyses was considered in situations where there were at least 3 treatment alternatives. When there were 2 treatment alternatives, pairwise meta-analysis was used to compare interventions. No network meta-analyses were carried out as part of the guideline development process.

Pairwise meta-analysis

Pairwise meta-analyses were performed in Cochrane Review Manager V5.3. 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 $I2 \geq 50\%$.

However, in cases where the results from individual pre-specified subgroup analyses were less heterogeneous (with $I2 < 50\%$) 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.

Data synthesis for qualitative studies

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, themes relating to information and misinformation) 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'. These review findings were evaluated using CERQual to examine their relevance to the review question, the importance given to each theme, and the extent to which each theme recurred across the different studies. The 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 studies and reviews

Where mixed methods studies were included in the reviews, the data was extracted and analysed separately for the quantitative and qualitative components using the relevant methods for each type of data (see above). If the data could not be analysed separately then no data synthesis was attempted, and the results were presented to the committee for discussion as individual studies. Any correlations or discrepancies between the findings of the mixed methods studies and the syntheses of the quantitative and qualitative findings of the above analyses will be noted.

Data synthesis for mixed methods sections of NICE evidence reviews

The quantitative and qualitative results were presented as a concept diagram with quantitative findings mapped onto the qualitative ones.

To do this the following approach was taken using the education interventions review as an example:

- A mixed methods summary diagram was produced which combined the main education-related findings from the qualitative barriers and facilitators review (evidence review B) with the relevant quantitative results from this review.
- Findings relating to infrastructure, were identified from review B and the ones that were considered to be most important were summarised in the summary of the evidence section within the education review chapter.
- These findings spanned the age groups and life stages and were further summarised to produce a diagram with key barriers and facilitators to vaccine uptake that related to education.
- Where possible links were made between barriers and corresponding facilitators that had been raised in the findings themselves or that were logically linked. So, for example, if a barrier concerned literacy problems and there was quantitative evidence from a study using video information about vaccines then the results of this study were summarised and placed in a box linked to the relevant barrier or facilitator.
- The quantitative evidence was then mapped onto the qualitative evidence.
- If a study could not be linked to a barrier or facilitator then it was shown in separate box at the side of the diagram.

Appraising the quality of evidence

Intervention studies

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 or uncontrolled before and after studies) were assessed using the preferred option specified in the NICE guidelines manual 2018 (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 clinical decision thresholds

The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to identify published minimal clinically important difference thresholds relevant to this guideline that might aid the committee in identifying clinical decision thresholds for the purpose of GRADE. No relevant MIDs were identified from the database. In addition, the Guideline Committee were asked to prospectively specify any outcomes where they felt a consensus clinical decision threshold could be defined from their experience. They were unable to define MIDs for the reviews in this guideline because they thought the clinically meaningful change in the outcomes (e.g., offers of vaccination, uptake of vaccines) may differ between vaccinations. Therefore, the line of no effect was used to downgrade for imprecision using GRADE.

GRADE for intervention studies analysed using pairwise analysis

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 3.

Table 3 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.
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 I2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I2 was less than 33.3%, the outcome was not downgraded. Serious: If the I2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I2 was greater than 66.7%, the outcome was downgraded two levels.
Imprecision	Where the line of no effect was defined as an MID for the outcome, it 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), and twice if the sample size of the study was sufficiently small that it is not plausible any realistic effect size could have been detected.
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

GRADE criteria	Reasons for downgrading quality
	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 'not 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.

Qualitative studies

Individual qualitative studies were quality assessed using the CASP qualitative checklist. Each individual study was classified into one of the following three groups:

- Low risk of bias – The findings and themes identified in the study are likely to accurately capture the true picture.
- Moderate risk of bias – There is a possibility the findings and themes identified in the study are not a complete representation of the true picture.
- High risk of bias – It is likely the findings and themes identified in the study are not a complete representation of the true picture

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

- Highly relevant – No important deviations from the protocol in perspective, population, phenomenon of interest and/or setting.
- Relevant – Important deviations from the protocol in one of perspective, population, phenomenon of interest and/or setting.
- Partially relevant – Important deviations from the protocol in at least two of the perspective, population, phenomenon of interest and/or setting.

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 then downgraded from this initial point as detailed in Table 4 below.

Table 4 Rationale for downgrading confidence in evidence for qualitative questions

CERQual criteria	Reasons for downgrading confidence
Methodological limitations	Not serious: If the theme was identified in studies at low risk of bias, the outcome was not downgraded

CERQual criteria	Reasons for downgrading confidence
	Serious: If the theme was identified only in studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If the theme was identified only in studies at high risk of bias, the outcome was downgraded two levels.
Relevance	High: If the theme was identified in highly relevant studies, the outcome was not downgraded Moderate: If the theme was identified only in in relevant and partially relevant studies, the outcome was downgraded one level. Low: If the theme was identified only in partially relevant studies, the outcome was downgraded two levels.
Coherence	Coherence was addressed based on two factors: Between study – does the theme represent the range of viewpoints covered by the included studies. Theoretical – does the theme provide a convincing theoretical explanation for the patterns found in the data. The outcome was downgraded once if there were concerns about one of these elements of coherence, and twice if there were concerns about both elements.
Adequacy of data	The outcome was downgraded if there was insufficient data to develop an understanding of the phenomenon of interest, either due to insufficient studies, participants or observations.

Mixed methods studies

Where published mixed methods studies were included in this guideline the following processes were followed:

- the quantitative and qualitative components had to both be relevant for inclusion using the same criteria for these types of evidence as detailed above and in the review protocols, otherwise only the relevant component was used if it could be extracted separately.
- if the data could be extracted separately, then:
    - the quantitative and qualitative components of the study were assessed for risk of bias using the relevant checklists for these study types (as detailed above) where possible.
    - the quantitative and qualitative results were extracted separately and analysed with other quantitative and qualitative data (as detailed above).
    - the quality of the synthesised evidence was assessed using GRADE or GRADE-CerQual as relevant (as detailed above above).
- if the quantitative and qualitative components were both relevant for inclusion but the data could not be extracted separately, then the quality of the study was assessed using the mixed methods appraisal tool (2018 version). The quality of the results was not assessed separately but covered as part of committee discussions about this evidence.

Reviewing economic evidence

Inclusion and exclusion of economic studies

Literature reviews seeking to identify published cost–utility analyses of relevance to the issues under consideration were conducted for all questions. In each case, the search undertaken for the clinical 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 clinical search. Initially only cost–utility analyses were included, but due to a lack of evidence in the younger age groups the criteria was expanded in those groups to include cost-effectiveness analyses. 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; 2014). This checklist is not intended to judge the quality of a study per se, but to determine whether an existing economic evaluation is useful to inform the decision-making of the committee for a specific topic within the guideline.

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 5.

Table 5 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 6.

Table 6 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 clinical evidence.

Costing exercises

To support recommendations with a potential resource impact where there was clinical evidence, simple cost-effectiveness analyses were conducted where the cost per additional person vaccinated was estimated. Cost-utility analyses were not conducted because there was not sufficient published evidence to inform the downstream costs and consequences of both the uptake interventions and current practice. Specially, the costs of the vaccines themselves are not known (as they are based on confidentially agreed prices with the DHSC), and because at most vaccination timepoints, vaccinations against multiple diseases would be given, this means that to accurately estimate QALY gains all of these potentially preventable diseases would need to be modelled, which was deemed to be unrealistic. Additionally, the vaccinations that are routine in the UK have already been deemed cost-effective, so it was not felt appropriate to be reassessing those consultations as part of this guideline, as that is the remit of the JCVI, not NICE.

Because cost-utility analyses were not conducted, the usual willingness-to-pay threshold of £20,000 per QALY gained could not be used in these costing analyses. Additionally, it was noted that this usual NICE threshold is only relevant if the resources considered are an NHS opportunity cost (i.e. where resources could theoretically be moved around between different services/parts of the NHS). For the majority of this guideline, the recommendations do not represent such a situation, but rather one where vaccine providers are trying to make the most efficient use of a finite set of resources available to them (and therefore opportunity costs fall more directly on the vaccine providers themselves, affecting other potential uptake increasing interventions that could be done).

This different situation therefore led to the committee interpreting these cost-effectiveness results in the following way. First, they agreed it was appropriate to be comparing different possible ways of increasing vaccine uptake, to identify the most cost-effective (i.e. something with a lower cost per additional person vaccinated should be given a higher priority for implementation than something with a higher cost per person vaccinated). Second, the committee agreed that it was not possible to draw a fixed threshold above which things would not be cost-effective. Rather, the guideline should recommend a range of possible interventions to increase uptake, with providers then implementing as many of those as possible within their resource constraints (again, starting with the most cost-effective). Finally, the committee agreed these cost-effectiveness results needed to be considered in combination with the other evidence in the guideline (such as the qualitative findings) when drawing conclusions, as there may be reasons (for example, equality issues) where simply adopting the approach that maximises the number of people vaccinated is not the correct choice, if for example it leads to substantial variations in uptake, or particular groups being poorly served. It was also noted that this is a health as well as an

equality issue, as the existence of particular communities/areas with low uptake rates can increase the risk of disease outbreaks, even if overall vaccination rates are high.

References

Follmann D, Elliott P, Suh I, Cutler J (1992) Variance imputation for overviews of clinical trials with continuous response. Journal of Clinical Epidemiology 45:769–73

Fu R, Vandermeer BW, Shamliyan TA, et al. (2013) Handling Continuous Outcomes in Quantitative Synthesis In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008. Available from http://www.ncbi.nlm.nih.gov/books/NBK154408/

Norman G., Sloan JA., Wyrwich KW. (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 41(5):582-92.