National Institute for Health and Care Excellence

Draft for consultation

Harmful gambling: identification, assessment and management

Methods

NICE guideline number tbc

Supplement 1

October 2023

Draft for Consultation

NICE accredited
www.nice.org.uk/accreditation

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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

© NICE, 2023. All rights reserved. Subject to Notice of rights.

ISBN:

Contents

Development of the guideline

Remit

The National Institute for Health and Care Excellence (NICE) commissioned the National Guideline Alliance (NGA) to develop a new guideline on harmful gambling. This guideline focuses on the identification, assessment and management of harmful gambling. In April 2022 the NGA became part of the NICE centre for guidelines.

To see "What this guideline covers" and "What this guideline does not cover" please see the guideline scope.

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

The review questions were based on the following frameworks:

- population, intervention, comparator and outcome (PICO) for reviews of interventions
- diagnostic reviews and reviews of prediction model accuracy – using population, diagnostic test (index test), reference standard and target condition (PIRT)
- qualitative reviews – using population, phenomenon of interest and context (PICo)

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

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

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

Evidence review	Review question	Type of review
[A] Factors suggesting harmful gambling	What factors, either alone or in combination, suggest that a person is participating in harmful gambling?	Diagnostic
[B] Tools for identification and assessment of harmful gambling	What is the accuracy of individual brief screening tools in identifying harmful gambling? What is the accuracy of tools to identify and assess harmful gambling?	Diagnostic
[C] Information and support	What are the information and support needs of people who participate in harmful gambling, their families, friends and others close to them?	Qualitative
[D] Models of care and service delivery	What is the effectiveness of different models of care and delivery of services for people who participate in harmful	Intervention

Evidence review	Review question	Type of review
	gambling (including those with comorbid conditions)?	
[E] Pharmacological treatment of harmful gambling	What is the effectiveness of pharmacological interventions for people who participate in harmful gambling (including those with comorbid conditions)?	Intervention
[F] Psychological and psychosocial treatment of harmful gambling	What is the effectiveness of psychological and psychosocial interventions for people who participate in harmful gambling (including those with comorbid conditions)?	Intervention[1]
[G] Interventions for families and affected others	What is the effectiveness of interventions and approaches for reducing gambling-related harms for families, friends and others close to people who gamble?	Intervention
[H] Relapse prevention	What is the effectiveness of interventions and approaches (for example, building recovery capital, mutual aid, peer support and mentoring programmes) for preventing relapse in people who have previously participated in harmful gambling?	Intervention
[I] Access	What are the barriers and facilitators to accessing treatment for harmful gambling from the perspective of practitioners, people who participate or have participated in harmful gambling, and their families, friends and others close to them?	Qualitative
[J] Interventions to improve access	What is the effectiveness of interventions or approaches designed to improve access to treatment for people who participate in harmful gambling among groups who are generally under-represented in treatment services?	Intervention
[K] Improving gambling	What works well and what could be improved in gambling treatment services, including	Qualitative

Evidence review	Review question	Type of review
treatment services	treatments for individuals, family approaches and relapse prevention, from the perspective of practitioners, people who participate or have participated in harmful gambling, and their families, friends and others close to them?	

[1]*Original health economic analysis conducted*

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

Searching for evidence

Scoping search

During the scoping phase, searches were conducted for previous guidelines, economic evaluations, health technology assessments, systematic reviews, randomised controlled trials, observational studies and qualitative research. Searches of websites of organisations, institutional repositories and internet search engines were also undertaken for relevant policies and related documents, including grey literature.

Systematic literature search

Systematic literature searches were undertaken to identify published evidence relevant to each review question. NICE undertook the searches for evidence reviews A-D and F-K, as reported below. For evidence review E, the Cochrane review for pharmacological interventions for the treatment of disordered and problem gambling (Dowling 2022) was used. A single search was used to cover the questions in evidence review F (psychological treatment) and evidence review H (relapse prevention).

Databases were searched using subject headings, free-text terms and, where appropriate, study type filters. Where possible, searches were limited to exclude: articles published before 2000; non-English language articles; non-human studies; and letters, editorials, case reports, conference proceedings and dissertations.

All the searches were conducted in the following databases: Applied Social Science Index and Abstracts (ASSIA), Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Embase, Emcare, Epistemonikos, Health Management Information Consortium (HMIC), International Health Technology Assessment Database (INAHTA), MEDLINE ALL, PsycInfo, Social Care Online (SCO), Social Policy & Practice (SPP) and Social Science Citation Index (SSCI). Searches for grey literature were conducted on the websites of relevant organisations (as listed in the protocols provided in Appendix A of each evidence review) for all review questions.

Searches were run once for evidence reviews A, B, C, I and K during development in 2022. Searches for evidence review F were updated in November 2022. Searches for evidence reviews D, G, H and J were updated in April 2023, 11 weeks in advance of the final committee meeting.

The Cochrane review (Dowling 2022) used for evidence review E searched the Cochrane Common Mental Disorders Specialised Register, CCTR, MEDLINE, Embase, and PsycInfo up to 11 January 2022 and the search strategies are provided in the appendix to the Cochrane review.

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

Economic systematic literature search

Systematic literature searches were also undertaken to identify published economic evidence. A combined literature search was undertaken to cover the economics aspects of all the review questions in a single search. Databases were searched using subject headings, free-text terms and, where appropriate, an economic evaluations search filter. Where possible, searches were limited to exclude: articles published before 2000; non-English language articles; non-human studies; and letters, editorials, case reports, conference proceedings and dissertations.

A search, based on the population search terms used in the evidence reviews, was conducted to identify economic evidence in the NHS Economic Evaluation Database (NHS EED) and INAHTA. Another search, using the population search terms based on the evidence reviews combined with an economic evaluations search filter, was conducted in ASSIA, CCTR, CINAHL, Embase, Emcare, HMIC, MEDLINE ALL, PsycInfo, SCO, SPP and SSCI. Where possible, searches were limited to studies published in English. Database strategies were limited to the publication dates of 2000 to current. Searches for grey literature were conducted on the websites of relevant organisations (as listed in the protocol provided in appendix A of each evidence review).

The economic literature searches were run in March 2022 and updated in April 2023, 11 weeks in advance of the final committee meeting before consultation on the draft guideline.

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

Quality assurance

Search strategies were quality assured by cross-checking reference lists of relevant studies, analysing search strategies from published systematic reviews and asking members of the committee to highlight key studies. The principal search strategies for each search were also quality assured by a second information scientist using an adaptation of the PRESS 2015 Guideline Evidence-Based Checklist (McGowan 2016).

Reviewing research evidence

Systematic review process

Evidence review E was based on a Cochrane systematic review, which the committee agreed answered the guideline review question about pharmacological interventions for the treatment of harmful gambling. No additional outcomes were of interest to the committee so no additional analysis was conducted by the NICE reviewers.

The evidence for all reviews except review E was reviewed in accordance with the following approach.

- Potentially relevant articles were identified from the search results for each review question by screening titles and abstracts. Full-text copies of the articles were then obtained.
- Full-text articles were reviewed against pre-specified inclusion and exclusion criteria in the review protocol (see Appendix A of each evidence review).
- Key information was extracted from each article on study methods and results, in accordance with factors specified in the review protocol. The information was presented in a summary table in the corresponding evidence review and in a more detailed evidence table (see Appendix D of each evidence review).
- Where data allowed, it was stratified upfront, according to the role of the gambling industry in funding the included studies (any industry funding/no industry funding/ unclear funding source). In the qualitative reviews, themes identified from industry funded evidence were therefore presented separately and in quantitative reviews, meta-analysis pooled results only from studies in the same funding stratification. This was in recognition that in this topic area of gambling, there are commercial interests about which the committee wanted to be aware and which they wanted to mitigate appropriately to reduce the risk of bias. The exception to this approach was review F in which the potential for funding related bias was explored through sensitivity analysis, which is explained below and in the review itself.
- Included studies were critically appraised using an appropriate checklist as specified in Developing NICE guidelines: the manual. Further detail on appraisal of the evidence is provided below.
- Summaries of quantitative evidence by outcome and qualitative evidence by theme were presented in the corresponding evidence review and discussed by the committee.

All review questions were subject to dual screening and study selection through a 10% random sample of articles. Any discrepancies were resolved by discussion between the first and second reviewers or by reference to a third (senior) reviewer. For review F, informing the network meta-analysis (NMA), dual data extraction was performed. For the remaining review questions, internal (NGA) quality assurance processes included consideration of the outcomes of screening, study selection and data extraction and the committee reviewed the results of study selection and data extraction. The review protocol for each question specifies whether dual screening and study selection was undertaken for that particular question. Drafts of all evidence reviews were quality assured by a senior reviewer.

Type of studies and inclusion/exclusion criteria

Inclusion and exclusion of studies was based on criteria specified in the corresponding review protocol. A general rule across reviews was that if some, but not all, of a study's participants were eligible for the review, then the study would be included if at least 66% of its participants met the protocol criteria.

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

For intervention reviews, randomised controlled trials (RCTs) were prioritised for inclusion because they are considered to be the most robust type of study design that could produce an unbiased estimate of intervention effects. In addition, experimental studies using a non-randomly assigned control group design that adjusted for relevant confounders or matched participants on important confounding domains were also considered for inclusion.

For diagnostic reviews, individual studies of diagnostic test accuracy were prioritised for inclusion. In addition, any study with random or consecutive selection of the target participants from which diagnostic data could be extracted. These were primarily cross-sectional or cohort studies but randomised controlled trials with one or more arms that met the inclusion criteria were also considered for inclusion.

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

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

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

Methods of combining evidence

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

Data synthesis for intervention studies

Pairwise meta-analysis

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

For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a fixed effect model was used to calculate risk ratios (RRs.

For continuous outcomes, measures of central tendency (mean) and variation (standard deviation; SD) are required for meta-analysis. Data for continuous outcomes, such as quality of life, were meta-analysed using an inverse-variance method for pooling weighted mean differences (WMDs). Where SDs were not reported for each intervention group, the standard error (SE) of the mean difference was calculated from other reported statistics (p values or 95% confidence intervals; CIs) and then meta-analysis was conducted as described above. The exception was the pairwise analysis in review F, where data for continuous outcomes were meta-analysed using random effects models of standardised mean differences (SMDs). A random effects model was used due to assumed heterogeneity based on differences between interventions that formed a class and methodological variation between studies, such as different tools for the measurement of the same outcomes. This was also in line with the approach taken to the NMA informing the same evidence review.

If a study reported only the summary statistic and 95% CI the generic-inverse variance method was used to enter data into RevMan5. If the control event rate was reported this was used to generate the absolute risk difference in GRADEpro. If multivariable analysis was used to derive the summary statistic but no adjusted control event rate was reported, no absolute risk difference was calculated. Where a study reported multiple adjusted estimates for the same outcome, the one that minimised the risk of bias due to confounding was chosen.

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

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

Network meta-analysis

Network meta-analysis (NMA) is a generalization of standard pairwise meta-analysis for A versus B trials, to data structures that include, for example, A versus B, B versus C, and A versus C trials (Dias 2011a; Lu 2004). A basic assumption of NMA methods is that direct and indirect evidence estimate the same parameter, that is, the relative effect between A and B measured directly from an A versus B trial, is the same with the relative effect between A and B estimated indirectly from A versus C and B versus C trials. NMA techniques strengthen inference concerning the relative effect of two treatments by including both direct and indirect comparisons between treatments, and, at the same time, allow simultaneous inference on all treatments examined in the pair-wise trial comparisons, which is essential for consideration of treatment in economic analysis (Caldwell 2005; Lu 2004). Simultaneous inference on the relative effect of a number of treatments is possible provided that treatments participate in a single “network of evidence”, that is, every treatment is linked to at least one of the other treatments under assessment through direct or indirect

comparisons. NMA takes all trial information into consideration, without ignoring part of the evidence and without introducing bias by breaking the rules of randomisation.

A key assumption when conducting an NMA is that the populations included in all randomised controlled trials (RCTs) considered in the NMA are similar so that the treatment effects are exchangeable across all populations (Mavridis 2015). This assumption of 'transitivity' of the effect may not hold if there are different potential effect modifiers that are not equally distributed across the different comparisons (Jansen 2014).

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

Due to the large number of interventions included in this review, comparing all pairs of interventions individually within the NMA would require multiple comparisons and complex consideration and interpretation of the evidence. Moreover, some interventions included in the systematic review had been tested on small numbers of participants and their effects were characterised by considerable uncertainty. For these reasons, the NMAs utilised class models: each class consisted of interventions with similar mode of action or similar treatment components or approaches, so that interventions within a class were expected to have similar (but not necessarily identical) effects. Use of class models in the NMA has three benefits:

- strength can be borrowed across interventions in the same class, therefore improving precision of effects
- networks that may otherwise be disconnected are possible to connect via interventions belonging to the same class, resulting in a connected network that includes all classes and interventions of interest
- relative effects between a more limited number of classes are easier to interpret and thus more helpful for the committee when making recommendations.

For all outcomes, both fixed and random class effects models were fitted. The random class effects model assumes the relative effects of treatments within a class are exchangeable. Treatment effects are shrunk towards a class mean and can borrow strength from other elements of the class. The fixed class effects model assumes treatments within a class have identical relative effects.

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

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

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

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

Overall methods and approaches adopted for the guideline NMA work were based on methodology described in the NICE Decision Support Unit (DSU) technical support document number 2 (Dias 2011a).

Details of the NMA methods employed in this guideline are provided in evidence review F (appendix L).

Data synthesis for diagnostic test accuracy reviews

When diagnostic test accuracy was measured dichotomously, sensitivity, specificity, and positive and negative predictive values were used as outcomes. These diagnostic test accuracy parameters were obtained directly from results reported in the source articles or calculated by the NGA technical team using data reported in the articles. Where possible, 95% CIs for diagnostic test accuracy parameters were reported; alternatively, median values and corresponding ranges were used if CIs were not reported and could not be calculated by the NGA technical team.

Meta-analysis of diagnostic test accuracy parameters would have been conducted if there were data from three or more studies that could be pooled but this did not occur.

Data synthesis for qualitative reviews

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

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

answering the review question, were also extracted and the same approach to 'thematic saturation' would have been applied. Thematic saturation was not reached for any themes in any of the qualitative reviews in this guideline. Therefore, all relevant data from all included qualitative studies were extracted and analysed.

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

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

Appraising the quality of evidence

Intervention studies

Pairwise meta-analysis

GRADE methodology for intervention reviews

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

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

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

The initial quality rating was based on the study design: RCTs and NRS assessed by ROBINS-I start as 'high' quality evidence, other non-randomised studies start as 'low' quality evidence. The rating was then modified according to the assessment of each quality element (Table 2). Each quality element considered to have a 'serious' or 'very serious' quality issue was downgraded by 1 or 2 levels respectively (for example, evidence starting as 'high' quality was downgraded to 'moderate' or 'low'

quality). In addition, there was a possibility to upgrade evidence from non-randomised studies (provided the evidence for that outcome had not previously been downgraded) if there was a large magnitude of effect, a dose–response gradient, or if all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results showed no effect.

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

Quality element	Description
Risk of bias ('Study limitations')	This refers to limitations in study design or implementation that reduce the internal validity of the evidence
Inconsistency	This refers to unexplained heterogeneity in the results
Indirectness	This refers to differences in study populations, interventions, comparators or outcomes between the available evidence and inclusion criteria specified in the review protocol
Imprecision	This occurs when a study has few participants or few events of interest, resulting in wide confidence intervals that cross minimally important thresholds
Publication bias	This refers to systematic under- or over-estimation of the underlying benefit or harm resulting from selective publication of study results

Table 3: GRADE quality ratings (by quality element)

Quality issues	Description
None or not serious	No serious issues with the evidence for the quality element under consideration
Serious	Issues with the evidence sufficient to downgrade by 1 level for the quality element under consideration
Very serious	Issues with the evidence sufficient to downgrade by 2 levels for the quality element under consideration

Table 4: Overall quality of the evidence in GRADE (by outcome)

Overall quality grading	Description
High	Further research is very unlikely to change the level of confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on the level of confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on the level of confidence in the estimate of effect and is likely to change the estimate
Very low	The estimate of effect is very uncertain

Assessing risk of bias in intervention reviews

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

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

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

- risk of bias arising from the randomization process
- risk of bias due to deviations from the intended interventions
- risk of bias due to missing outcome data
- risk of bias due to measurement of the outcome
- risk of bias in selection of the reported result.

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

More details about the Cochrane risk of bias tool can be found in Section 8 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011).

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

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

Assessing inconsistency in intervention reviews

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

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

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

Assessing indirectness in intervention reviews

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

Assessing imprecision and importance in intervention reviews

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

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

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

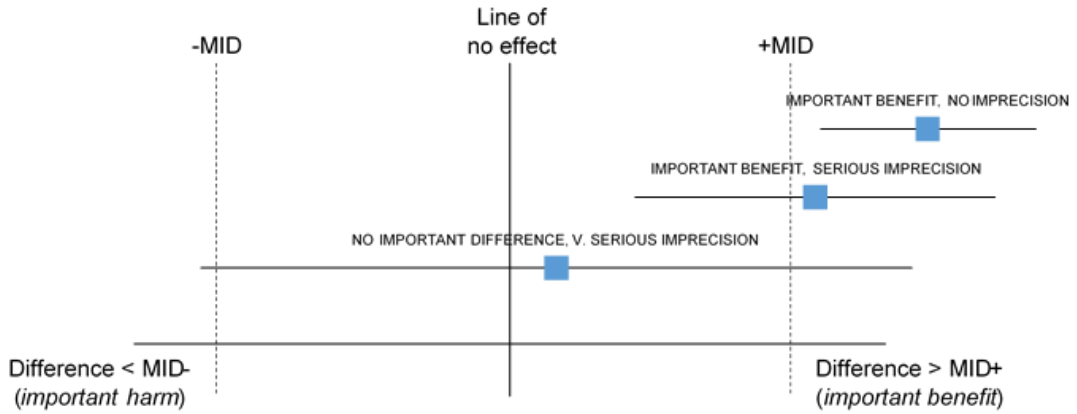
When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

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

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

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

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



MID, minimally important difference

Defining minimally important differences for intervention reviews

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

In the absence of published or accepted MIDs, the committee agreed to use the GRADE default MIDs to assess imprecision. For dichotomous outcomes minimally important thresholds for a RR of 0.8 and 1.25 respectively were used as default MIDs in the guideline. The committee also chose to use 0.8 and 1.25 as the MIDs for ORs & HRs in the absence of published or accepted MIDs. ORs were predominantly used in the guideline when Peto OR were indicated due to low event rates, at low event rates OR are mathematically similar to RR making the extrapolation appropriate. While no default MIDs exist for HR, the committee agreed for consistency to continue to use 0.8 and 1.25 for these outcomes. For continuous outcomes measured using SMD, minimally important thresholds of -0.5 and 0.5 respectively were used as default MIDs in the guideline.

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

Assessing publication bias in intervention reviews

Where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias. Where fewer than 10 studies were included for an outcome, the committee subjectively assessed the likelihood of publication bias based on factors such as the proportion of trials funded by industry and the propensity for publication bias in the topic area.

Network meta-analysis

For the NMAs, quality was assessed by looking at risk of bias across the included evidence using the Cochrane Risk of Bias Tool for Randomized Controlled Trials, as

well as heterogeneity and consistency (also called coherence). Heterogeneity refers to the differences in treatment effects between trials within each treatment contrast (measured by the posterior median between-study standard deviation and compared with treatment posterior mean effects), while consistency refers to the differences between the direct and indirect evidence informing the treatment contrasts. Direct and indirect comparisons measure the same underlying true effect, and therefore, in principle, they should be consistent. However, this is not the case if effect modifiers and heterogeneity across studies, populations and comparisons are present. Inconsistency arises when there is a conflict between direct evidence (from an A vs. B trial) and indirect evidence (gained from A vs. C and B vs. C trials) and can only be assessed when there are both direct and indirect sources of evidence for a treatment comparison, that is, there are closed loops of evidence on three treatments that are informed by at least three distinct trials (Caldwell 2014).

Checking for inconsistency between direct and indirect evidence can reveal whether the transitivity assumption holds. To determine if there was evidence of inconsistency, in each analysis, the selected consistency model (fixed or random effects) was compared to an “inconsistency”, or unrelated mean effects, model (Dias 2011b & 2013). When evidence of inconsistency was found, studies contributing to between-trial heterogeneity were checked for data accuracy and analyses were repeated if corrections in the data extraction were made. However, following any data corrections and if inconsistency persisted, no studies were excluded from the analysis, as their results could not be considered as less valid than those of other studies solely because of the inconsistency findings. Nevertheless, the presence of inconsistency in the network was highlighted and results were interpreted accordingly by the committee.

Tests of inconsistency are inherently underpowered, so they may fail to detect inconsistency even though this may be present in the network (Dias 2011b). Therefore, even if inconsistency is not detected, results of NMA should be interpreted following qualitative evaluation of the anticipated transitivity within the network and judgement of reasons for potential inconsistency (Linde 2016).

Publication bias is known to affect results of meta-analyses (Moreno 2009 & 2011; Turner 2008). Small sample size studies are associated with publication bias as small studies with positive results are more likely to be published compared with small studies with negative results, and may also be associated with lower study quality. Published smaller studies tend to overestimate the relative treatment effect of interventions versus control, compared to larger studies (Chaimani 2013; Moreno 2011). Bias adjustment models were fitted to down-weight trials with industry or unclear funding. Models that adjusted for small study bias were also fitted (Dias 2010, Welton 2009).

Threshold analysis was done to test the robustness of treatment recommendations based on the NMA, to potential biases or sampling variation in the included evidence. Threshold analysis has been developed as an alternative to GRADE for assessing confidence in guideline recommendations based on network meta-analysis (Phillippo 2019).

Diagnostic studies

Adapted GRADE methodology for diagnostic reviews

For diagnostic reviews, an adapted GRADE approach was used. GRADE methodology is designed for intervention reviews but the quality assessment elements and outcome presentation were adapted by the guideline developers for diagnostic test accuracy reviews. For example, GRADE tables were modified to include diagnostic test accuracy measures (sensitivity, specificity and predictive values).

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

The initial quality rating was based on the study design: cross-sectional or cohort studies start as 'high' quality and case–control studies start as 'low' quality.

Table 5: Adaptation of GRADE quality elements for diagnostic reviews

Quality element	Description
Risk of bias ('Study limitations')	Limitations in study design and implementation may bias estimates of diagnostic accuracy. High risk of bias for the majority of the evidence reduces confidence in the estimated effect. Diagnostic accuracy studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)
Inconsistency	This refers to unexplained heterogeneity in test accuracy measures (such as sensitivity and specificity) between studies
Indirectness	This refers to differences in study populations, index tests, reference standards or outcomes between the available evidence and inclusion criteria specified in the review protocol
Imprecision	This occurs when a study has relatively few participants and the probability of a correct diagnosis is low. Accuracy measures would therefore have wide confidence intervals around the estimated effect

Assessing risk of bias in diagnostic reviews and prediction models

Risk of bias in diagnostic reviews and prediction models was assessed using the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) checklist (see Appendix H in Developing NICE guidelines: the manual).

Risk of bias in primary diagnostic accuracy reviews or prediction models in QUADAS-2 consists of 4 domains:

- participant selection
- index test
- reference standard
- flow and timing.

More details about the QUADAS-2 tool can be found on the developer's website.

Assessing inconsistency in diagnostic reviews

Inconsistency refers to the unexplained heterogeneity of the results in meta-analysis. When estimates of diagnostic accuracy and prediction model parameters vary widely across studies (that is, there is heterogeneity or variability in results), this suggests true differences in underlying effects. Inconsistency is, thus, only truly applicable when statistical meta-analysis is conducted (that is, results from different studies are pooled).

Inconsistency for diagnostic reviews and prediction models was assessed based on visual inspection of the point estimates and confidence intervals of the included studies. If these varied widely (for example, point estimates for some studies lying outside the CIs of other studies) the evidence was downgraded for inconsistency.

Assessing indirectness in diagnostic reviews

Indirectness in diagnostic reviews and prediction models was assessed using the QUADAS-2 checklist by assessing the applicability of the studies in relation to the review question in the following domains:

- participant selection
- index test
- reference standard.

More details about the QUADAS-2 tool can be found on the developer's website.

Assessing imprecision and importance in diagnostic reviews

The judgement of precision for diagnostic evidence was based on the CIs of the single pair of parameters prioritised by the guideline committee. For review A, the judgement of precision was based on positive predictive values and for review B, it was based on sensitivity and specificity.

For review A, the committee agreed the following cut-off when summarising the performance of factors suggesting harmful gambling in terms of positive predictive values (PPVs), for studies where people presented in a non-gambling specialist setting. This was the value above which the risk factor should be considered an indication of harmful gambling behaviour:

- important factor: PPV 2% or greater.

This was used as the upper threshold to asses the imprecision of PPV and the committee set a lower threshold of 0.5%, based on the prevalence figure for 'problem' gambling in England (PHE, 2022).

Outcomes were downgraded for imprecision when their 95% CI crossed at least 1 threshold. If the CI crossed 1 threshold, the outcome was downgraded one level for imprecision. If the CI crossed 2 thresholds, the outcome was downgraded two levels for imprecision. These assessments were made on the meta-analysed outcomes where applicable or, if outcomes were not meta-analysed, on the individual study results themselves. For negative predictive values (NPVs), the committee chose thresholds of 98% and 99.5% and they were used to assess imprecision in the same way.

For sensitivity and specificity the committee defined 2 decision thresholds, a value above which the measurement tools could be recommended and a value below which the tools would be considered of no use. These thresholds were based on the committee's experience and consensus.

The thresholds were:

- sensitivity: low threshold 60%, high threshold 90%
- specificity: low threshold 60%, high threshold 90%.

These thresholds were also used to assess imprecision of sensitivity and specificity.

Outcomes were downgraded for imprecision when their 95% CI crossed at least 1 threshold. If the CI crossed 1 threshold, the outcome was downgraded once for imprecision. If the CI crossed 2 thresholds, the outcome was downgraded twice for imprecision. These assessments were made on the meta-analysed outcomes where applicable or if outcomes were not meta-analysed, on the individual study results themselves.

Qualitative studies

GRADE-CERQual methodology for qualitative reviews

For qualitative reviews an adapted GRADE Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach (Lewin 2018) was used. In this approach the quality of evidence is considered according to themes in the evidence. The themes may have been identified in the primary studies or they may have been identified by considering the reports of a number of studies. Quality elements assessed using GRADE-CERQual are listed and defined in Table 6. Each element was graded using the levels of concern summarised in Table 7.

The ratings for each component were combined (as with other types of evidence) to obtain an overall assessment of quality for each theme as described in Table 8. 'Confidence' in this context refers to the extent to which the review finding is a reasonable representation of the phenomenon of interest set out in the protocol. Similar to other types of evidence all review findings start off with 'high confidence' and are rated down by one or more levels if there are concerns about any of the individual CERQual components. In line with advice from the CERQual developers, the overall assessment does not involve numerical scoring for each component but in order to ensure consistency across and between guidelines, the NGA established some guiding principles for overall ratings. For example, a review finding would not be downgraded (and therefore would be assessed with 'high' confidence) if at least 2 of the individual components were rated as 'no or very minor; and none of the components were rated as having moderate or serious concerns.

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

Table 6: Adaptation of GRADE quality elements for qualitative reviews

Quality element	Description
Methodological limitations	Limitations in study design and implementation may bias interpretation of qualitative themes identified. High risk of bias for the majority of the evidence reduces our confidence that the review findings reflect the phenomena of interest. Qualitative studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)
Relevance (or applicability) of evidence	This refers to the extent to which the context of the studies supporting the review findings is applicable to the context specified in the review question
Coherence of findings	This refers to the extent to which review findings are well grounded in data from the contributing primary studies and provide a credible explanation for patterns identified in the evidence. If the data from the underlying studies are ambiguous or contradict the review finding this would reduce our confidence in the finding.
Adequacy of data (theme saturation or sufficiency)	This corresponds to a similar concept in primary qualitative research, that is, whether a theoretical point of theme saturation was achieved, at which point no further citations or observations would provide more insight or suggest a different interpretation of the particular theme. Judgements are not based on the number of studies but do take account of the quantity and also richness of data underpinning a finding. The more complex the finding, the more detailed the supporting data need to be. For simple findings, relatively superficial data would be considered adequate to explain and explore the phenomenon being described.

Table 7: CERQual levels of concern (by quality element)

Level of concern	Definition
None or very minor concerns	Unlikely to reduce confidence in the review finding
Minor concerns	May reduce confidence in the review finding
Moderate concerns	Will probably reduce confidence in the review finding
Serious concerns	Very likely to reduce confidence in the review finding

Table 8: Overall confidence in the evidence in CERQual (by review finding)

Overall confidence level	Definition
High	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest
Moderate	It is likely that the review finding is a reasonable representation of the phenomenon of interest
Low	It is possible that the review finding is a reasonable representation of the phenomenon of interest
Very low	It is unclear whether the review finding is a reasonable representation of the phenomenon of interest

Assessing methodological limitations in qualitative reviews

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

Table 9: Methodological limitations in qualitative studies

Aim and appropriateness of qualitative evidence	This domain assesses whether the aims and relevance of the study were described clearly and whether qualitative research methods were appropriate for investigating the research question
Rigour in study design or validity of theoretical approach	This domain assesses whether the study approach was documented clearly and whether it was based on a theoretical framework (such as ethnography or grounded theory). This does not necessarily mean that the framework has to be stated explicitly, but a detailed description ensuring transparency and reproducibility should be provided
Sample selection	This domain assesses the background, the procedure and reasons for the method of selecting participants. The assessment should include consideration of any relationship between the researcher and the participants, and how this might have influenced the findings
Data collection	This domain assesses the documentation of the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses who conducted any interviews, how long they lasted and where they took place
Data analysis	This domain assesses whether sufficient detail was documented for the analytical process and whether it was in accordance with the theoretical approach. For example, if a thematic analysis was used, the assessment would focus on the description of the approach used to generate themes. Consideration of data saturation would also form part of this assessment (it could be reported directly or it might be inferred from the citations documented that more themes could be found)
Results	This domain assesses any reasoning accompanying reporting of results (for example, whether a theoretical proposal or framework is provided)

Assessing relevance of evidence in qualitative reviews

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

Assessing coherence of findings in qualitative reviews

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

Assessing adequacy of data in qualitative reviews

Adequacy of data (theme saturation or sufficiency) corresponds to a similar concept in primary qualitative research in which consideration is made of whether a theoretical point of theme saturation was achieved, meaning that no further citations or observations would provide more insight or suggest a different interpretation of the theme concerned. As noted above, it is not equivalent to the number of studies contributing to a theme, but it does take account of the quantity of data supporting a review finding (for instance whether sufficient quotations or observations were provided to underpin the findings) and in particular the degree of 'richness' of supporting data. Concerns about richness arise when insufficient details are provided by the data to enable an understanding of the phenomenon being described. Generally, if a review finding is fairly simple then relatively superficial data will be needed to understand it. Data underpinning a more complex finding would need to offer greater detail, allowing for interpretation and exploration of the phenomenon being described. Therefore in assessing adequacy our downgrading involved weighing up the complexity of the review finding against the explanatory contribution of the supporting data.

Reviewing economic evidence

Inclusion and exclusion of economic studies

Systematic reviews of economic literature were conducted in all areas covered in the guideline. Titles and abstracts of articles identified through the economic literature searches were assessed for inclusion using the predefined eligibility criteria listed in Table 10.

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

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

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

Details of economic evidence study selection, lists of included and excluded studies, summaries of economic evidence in economic evidence profiles, and economic evidence tables for each review question are presented in respective evidence reviews.

Appraising the quality of economic evidence

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

The methodological assessment of economic studies considered in this guideline has been summarised in economic evidence profiles that were developed for each review question for which economic evidence was available. All studies that fully or partially

met the applicability and quality criteria described in the methodology checklist were considered during the guideline development process.

Inclusion and exclusion of health state utility studies

Literature on the health-related quality of life of adults experiencing harmful gambling was systematically searched to identify studies reporting appropriate utility scores that could be utilised in a primary economic modelling. The titles and abstracts of papers identified through the searches were independently assessed for inclusion using predefined eligibility criteria defined in Table 11.

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

Inclusion criteria
Only studies from Organisation for Economic Co-operation and Development member countries were included, as the aim of the review was to identify utility data transferable to the UK context.
Studies should report utility data for health states associated with harmful gambling through the care pathway.
Studies should report health-related quality of life ratings made using a validated generic or harmful gambling-specific preference-based measure directly or via mapping from another validated non-preference-based measure. Utility values should have been elicited from the general population using a choice-based method, such as time trade-off (TTO) or standard gamble (SG).
Exclusion criteria
Poster presentations and abstracts in conference proceedings
Non-English language papers

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

Utility studies that met inclusion criteria and those that were excluded after full text was obtained are listed in evidence review F, which included economic modelling.

Economic modelling

The aims of the economic input to the guideline were to inform the guideline committee of potential economic issues to ensure that recommendations represented a cost effective use of healthcare resources. Economic evaluations aim to integrate data on healthcare benefits (ideally in terms of quality-adjusted life-years; QALYs) with the costs of different options. In addition, the economic input aimed to identify areas of high resource impact; these are recommendations which (while cost effective) might have a large impact on Clinical Commissioning Group or Trust finances and so need special attention.

Areas for economic modelling were prioritised by the committee. The rationale for prioritising review questions for economic modelling was set out in an economic plan agreed between NICE, the committee, and members of the NGA technical team. Economic modelling was undertaken in areas with likely major resource implications, where the current extent of uncertainty over cost effectiveness was significant and

economic analysis was expected to reduce this uncertainty. The following economic questions were selected as key issues to be addressed by economic modelling:

- Cost-effectiveness of pharmacological and psychological/psychosocial treatments for adults experiencing harmful gambling. As the size of the available clinical evidence for pharmacological interventions was small and characterised by limitations, it did not allow the development of a robust model. Instead, a simple cost analysis was undertaken, to estimate the intervention costs of naltrexone, which was the only pharmacological intervention considered by the committee for a recommendation. This is described in evidence review E. The methods and results of the de novo economic analysis for psychological/psychosocial treatments are fully reported in appendix I of evidence review F.
- Cost-effectiveness of interventions and approaches for preventing relapse in people who have previously participated in harmful gambling. No economic model was carried out for this question, due to the limited amount and quality of the clinical evidence, which did not allow for a robust model to be developed or for recommendations on specific interventions to be made.

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

Cost effectiveness criteria

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

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

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

Developing recommendations

Guideline recommendations

Recommendations were drafted on the basis of the committee's interpretation of the available evidence, taking account of the balance of benefits, harms and costs between different courses of action. When effectiveness, qualitative and economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on their expert opinion. The considerations for making

consensus-based recommendations include the balance between potential benefits and harms, the economic costs or implications compared with the economic benefits, current practices, recommendations made in other relevant guidelines, person's preferences and equality issues.

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

For further details refer to Developing NICE guidelines: the manual.

Research recommendations

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

Validation process

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

Updating the guideline

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

Funding

The NGA was commissioned by NICE to develop this guideline. In April 2022 the NGA became part of the NICE centre for guidelines.

References

Caldwell 2005

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

Caldwell 2014

Caldwell DM (2014). An overview of conducting systematic reviews with network meta-analysis. Syst Rev, 3:109.

Chaimani 2013

Chaimani A, Vasiliadis HS, Pandis N, Schmid CH, Welton NJ, Salanti G (2013). Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. International journal of epidemiology, 42(4), 1120-1131.

Dias 2010

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

Dias 2011a

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

Dias 2011b

Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE (2011b). NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials (last updated 2014) http://www.nicedsu.org.uk

Dias 2013

Dias, D., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., Ades, A.E (2013). Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Medical Decision Making, 33, 641-56.

Dixon-Woods 2005

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

Hayden 2013

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

Higgins 2011

Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated 2022] The Cochrane Collaboration. Available from www.handbook.cochrane.org (accessed 03 September 2023)

Jansen 2014

Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G (2014). Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health, 17(2), 157-73.

Lewin 2018

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

Linde 2016

Linde K, Rücker G, Schneider A, Kriston L (2016). Questionable assumptions hampered interpretation of a network meta-analysis of primary care depression treatments. J Clin Epidemiol, 71, 86-96., 2016.

Lu 2004

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

Mavridis 2015

Mavridis D, Giannatsi M, Cipriani A, Salanti G (2015). A primer on network meta-analysis with emphasis on mental health. Evid Based Ment Health, 18(2), 40-46.

McGowan 2016

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

Moreno 2009

Moreno SG, Sutton AJ, AdesA, Stanley TD, Abrams KR, Peters JL, Cooper NJ (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC medical research methodology, 9(1), 2.

Moreno 2011

Moreno SG, Sutton AJ, Ades A, Cooper NJ, Abrams KR (2011). Adjusting for publication biases across similar interventions performed well when compared with gold standard data. Journal of clinical epidemiology, 64(11), 1230-1241.

NICE 2018

National Institute for Health and Care Excellence (NICE) (2014) NICE Policy on conflicts of interest (updated 2017). Available from https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-procedures/declaration-of-interests-policy.pdf (accessed 03 September 2023)

Phillippo 2019

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

Public Health England 2022

Gambling related harms evidence review: summary (updated 2023) Available from Gambling-related harms evidence review: summary - GOV.UK (www.gov.uk) (accessed 09 June 2023)

Santesso 2016

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

Turner 2008

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. New England Journal of Medicine, 358(3), 252-260.

Welton 2009

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