National Institute for Health and Care Excellence

Draft for consultation

Self-harm: assessment, management and preventing recurrence

NICE guideline: methods

NICE guideline number tbc

Methods

January 2022

Draft for Consultation

Evidence reviews were developed by the National Guideline Alliance hosted by the Royal College of Obstetricians and Gynaecologists

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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

Contents

Development of the guideline

Remit

The National Institute for Health and Care Excellence (NICE) commissioned the National Guideline Alliance (NGA) to update the following 2 guidelines on self-harm:

- Self-harm in over 8s: short-term management and prevention of recurrence (CG16)
- Self-harm in over 8s: long-term management (GC133)

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 Developing NICE guidelines: the manual.

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

Developing the review questions and outcomes

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

The review questions were based on the following frameworks:

- population, intervention, comparator and outcome (PICO) for reviews of interventions
- qualitative reviews – using population, phenomenon of interest and context (PICo)

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

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

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

Evidence review	Review question	Type of review
A – information and support needs of people who have self-harmed	What are the information and support needs of people who have self-harmed?	Qualitative
B – information and support needs of families and carers of people who have self-harmed	What are the information and support needs of the families and carers of people who have self-harmed?	Qualitative
C – consent, confidentiality and safeguarding	What is the most effective approach to obtain consent, ensure confidentiality and promote safeguarding when people have self-harmed?	Intervention
D - involving family and carers in the management of people who have self-harmed	What are the views and preferences of people who have self-harmed, their families and carers, and staff working with people who have self-harmed about the best ways of involving family and carers in the management of people who have self-harmed?	Qualitative
E –assessment in specialist settings	How should assessment for people who have self-harmed be undertaken in specialist settings, such as: community mental health services, emergency	Intervention

Evidence review	Review question	Type of review
	departments (by specialist staff), inpatient mental health services?	
F – assessment in non-specialist settings	How should assessment for people who have self-harmed be undertaken in non-specialist settings, such as: primary care, social care, community pharmacy, ambulances, emergency departments (by non-specialist staff), schools, colleges and universities, the criminal justice system and immigration removal centres and acute general hospitals?	Intervention
G – risk assessment and formulation	What are the benefits and harms of a risk assessment and formulation including those models or tools that combine elements of machine learning and artificial intelligence for people who have self-harmed?	Intervention
H – admission to hospital	What are the benefits and harms associated with admission to acute general hospital for people who have self-harmed but no longer require physical care?	Intervention
I – initial after-care	How should initial after-care be provided to people following an episode of self-harm?	Intervention
J – psychological and psychosocial interventions	What psychological and psychosocial interventions (including safety plans and electronic health-based interventions) are effective for people who have self-harmed?	Intervention[1]
K – pharmacological interventions	What pharmacological interventions are effective for people who have self-harmed?	Intervention
L – harm minimisation strategies	What is the effectiveness of harm minimisation strategies for people who have self-harmed?	Intervention
M – therapeutic risk-taking strategies	What is the effectiveness of therapeutic risk-taking strategies for people who have self-harmed?	Intervention
N – supporting people to be safe after self-harm	What are the most effective ways of supporting people to be safe after self-harm?	Intervention
O – safer prescribing	What are the key principles of safer prescribing for people who have self-harmed?	Intervention
P – skills required for staff in specialist mental health settings who assess and treat people who have self-harmed	What are the views and preferences of staff in specialist mental health settings, people who have self-harmed and their family members/carers about what skills are required for staff in specialist mental health settings who assess and treat people who have self-harmed?	Qualitative
Q – supervision required for staff in specialist mental health settings who assess and treat	What are the views and preferences of staff in specialist mental health settings about what supervision is required for staff in specialist mental health settings who	Qualitative

Evidence review	Review question	Type of review
people who have self-harmed	assess and treat people who have self-harmed?	
R – skills required for staff in non-specialist settings who assess and treat people who have self-harmed	What are the views and preferences of staff in non-specialist settings, people who have self-harmed and their family members/carers about what skills are required for staff in non-specialist settings who assess and treat people who have self-harmed?	Qualitative
S – supervision required for staff in non-specialist mental health settings who assess and treat people who have self-harmed	What are the views and preferences of staff in non-specialist mental health settings about what supervision is required for staff in non-specialist mental health settings who assess and treat people who have self-harmed?	Qualitative
T models of care	What are the most effective models of care for people who have self-harmed?	Intervention

[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.

Additional information related to development of the guideline is contained in:

- Supplement 3 (NGA staff list).

Searching for evidence

Scoping search

During the scoping phase, searches were conducted for previous guidelines, systematic reviews, policy papers, economic evaluations and health technology assessments.

Systematic literature search

Systematic literature searches were undertaken to identify published evidence relevant to each review question.

Databases were searched using subject headings, free-text terms and, where appropriate, study type filters. Where possible, searches were limited to retrieve studies published in English. All the searches were conducted in the following databases: Embase, Medline, Medline-in-Process, Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), International Health Technology Assessments (IHTA) and PsycINFO. For review questions related to nursing, Emcare and CINAHL were also searched. For review questions where key papers were supplied pre-search, forward and backward citation searching was undertaken in the Web of Science along with checking the reference lists.

Searches were run once for all reviews during development. Searches for the following question were updated seven weeks in advance of the final committee meeting.

- H. What are the benefits and harms associated with admission to acute general hospital for people who have self-harmed but no longer require physical care?

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

Economic systematic literature search

Systematic literature searches were also undertaken to identify published economic evidence. Databases were searched using subject headings, free-text terms and, where appropriate, an economic evaluations search filter.

A single search, using the population search terms used in the evidence reviews, was conducted to identify economic evidence in the NHS Economic Evaluation Database (NHS EED) and IHTA. Another single search, using the population search terms used in the evidence reviews combined with an economic evaluations search filter, was conducted in Medline, Medline in Process, CCTR and Embase. Where possible, searches were limited to studies published in English.

As with the general literature searches, the economic literature searches were updated seven 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 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). In addition, all publications highlighted by stakeholders at the time of the consultation on the draft scope were considered for inclusion.

Reviewing research evidence

Systematic review process

When the guideline started development, the NGA was at the beginning of a phased transition from using STAR software to manage the evidence reviews to using EPPI Reviewer software. Moreover, EPPI Reviewer was also undergoing further development during the development of this guideline. As a consequence, the initial review conducted for the guideline (“H. What are the benefits and harms associated with admission to acute general hospital for people who have self-harmed but no longer require physical care?”) was undertaken in STAR with the subsequent reviews

undertaken in EPPI. Although the content of the reviews does not differ between STAR and EPPI Reviewer or between the different EPPI Reviewer updates, the presentation of the contents do in some cases in terms of style and formatting, for example for references, PRISMA diagram, evidence tables and excluded studies. The evidence was reviewed in accordance with the following approach.

- Potentially relevant articles were identified from the search results for each review question by screening titles and abstracts. Full-text copies of the articles were then obtained.
- Full-text articles were reviewed against pre-specified inclusion and exclusion criteria in the review protocol (see Appendix A of each evidence review).
- Key information was extracted from each article on study methods and results, in accordance with factors specified in the review protocol. The information was presented in a summary table in the corresponding evidence review and in a more detailed evidence table (see Appendix D of each evidence review).
- Included studies were critically appraised using an appropriate checklist as specified in Developing NICE guidelines: the manual. Further detail on appraisal of the evidence is provided below.
- Summaries of evidence by outcome 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. All the review questions were also subject to internal (NGA) quality assurance processes including consideration of the outcomes of screening, study selection and data extraction, and the committee reviewed the results of study selection and data extraction. Drafts of all evidence reviews were quality assured by a senior reviewer.

The process of study selection for review questions selected as high priorities for economic analysis (and those selected as medium priorities and where economic analysis could influence recommendations), were checked by a senior health economist.

Type of studies and inclusion/exclusion criteria

Inclusion and exclusion of studies was based on criteria specified in the corresponding review protocol.

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

For intervention reviews, randomised controlled trials (RCTs) were prioritised for inclusion because they are considered to be the most robust type of study design that could produce an unbiased estimate of intervention effects. Where there was limited evidence from RCTs, non-randomised studies (NRS) were 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, unpublished studies and studies published in languages other than English were excluded. Conference abstracts were not considered for inclusion because conference abstracts typically do not have sufficient information to allow for full critical appraisal.

Methods of combining evidence

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

Data synthesis for intervention studies

Pairwise meta-analysis

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

For dichotomous outcomes, such as rate of self-harm, the Mantel–Haenszel method with a fixed effect model was used to calculate risk ratios (RRs). For all outcomes with zero events in both arms the risk difference was presented. For outcomes in which the majority of studies had low event rates (<1%), Peto odds ratios (ORs) were calculated as this method performs well when events are rare (Bradburn 2007).

For continuous outcomes, measures of central tendency (mean) and variation (standard deviation; SD) are required for meta-analysis. 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.

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.

When evidence was based on studies that reported descriptive data or medians with interquartile ranges or p values, this information was included in the corresponding GRADE tables (see below) without calculating relative or absolute effects. Although effects were not included in the GRADE tables, these data were considered during committee discussions of the evidence.

For some reviews, evidence was either stratified from the outset or separated into subgroups when heterogeneity was encountered. The stratifications and potential subgroups were pre-defined at the protocol stage (see the protocols for each review for further detail). Where evidence was stratified or 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).

Included Cochrane Reviews

During the development of this guideline, two registered Cochrane protocols were identified which matched the committee's intended review questions:

- J – psychological and psychosocial interventions
- K – pharmacological interventions.

The Cochrane review team completed two reviews investigating the effectiveness of psychosocial interventions in adults (Witt 2021a) and psychosocial and pharmacological interventions in children and young people (CYP) (Witt 2021b) during guideline development and presented their results to the guideline committee, which used them to make recommendations.

Cochrane's methods are closely aligned to standard NICE methods, minor deviations (the use of GRADE only on main outcomes with no overall quality rating for those with zero events in either arm, summary of findings tables instead of full GRADE tables, defining primary and secondary outcomes as opposed to critical and important and including countries from a broader range of income categories than the majority of the other reviews in the guideline) relevant to the topic area were highlighted to the committee and taken into account in discussions of the evidence.

Data synthesis for qualitative reviews

Where possible, a meta-synthesis was conducted to combine evidence from qualitative studies. 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.

In qualitative synthesis, a theme being reported more than other themes across included studies does not necessarily mean that the theme is more important than other themes. The aim of qualitative research is to identify new perspectives on a particular topic. Study types and populations in qualitative research can differ widely, meaning that themes identified by just one or a few studies can provide important new information on a given topic. Therefore, for the purpose of the qualitative reviews in this guideline, it was planned that further studies would not be added when they reported the same themes as had already been identified from other UK-based studies because the emphasis was to be on conceptual robustness and relevance rather than quantitative completeness of the evidence.

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

Modified 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 a modified version of 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 certainty of evidence for 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 was not included in the GRADE table, but was considered during committee discussions of the evidence, taking into account 95% confidence intervals around the point estimate of the effect, any relevant MIDs, committee expertise and the effect of a single intervention based on multiple outcomes.
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 revised Cochrane risk of bias tool (RoB 2; see Appendix H in Developing NICE guidelines: the manual; NICE 2014).

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 AMSTAR checklist was used and for systematic reviews of other study types the ROBIS checklist was used (see Appendix H in Developing NICE guidelines: the manual; NICE 2014).

For non-randomised studies the ROBINS-I checklist was used (see Appendix H in Developing NICE guidelines: the manual; NICE 2014).

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

A modified version of the GRADE approach to rate the certainty of evidence in systematic reviews was used. The modification of the usual GRADE approach was part of a pilot project undertaken by NICE, to examine the assessment of certainty of evidence in systematic reviews. Instead of using predefined clinical decision/minimal important difference (MID) thresholds to assess imprecision in GRADE tables, imprecision was assessed qualitatively during committee discussions. These discussions involved consideration of published MIDs where they existed (see also next section), but the committee were also encouraged to make judgements of imprecision based on the 95% confidence intervals and sample sizes reported in the GRADE profiles. The committee were not aware of any published MIDs for any of the outcomes in the intervention reviews and so the discussions were based on the width of confidence intervals and whether they crossed the line of no effect. This should enable judgements of clinical importance to be made in the context of wider decision making, taking into account evidence across all outcomes and analyses, including health economic analyses.

Committee discussions regarding the clinical importance of effects was recorded in the 'imprecision and clinical importance of effects' section of the evidence review. In particular, this included consideration of whether the whole effect of a treatment (which may be felt across multiple independent outcome domains) would be likely to be clinically meaningful, rather than simply whether each individual sub outcome might be meaningful in isolation. The impact of imprecision on the recommendations was presented in the 'quality of the evidence' section of the committee discussion in the evidence review

Defining minimally important differences for intervention reviews

The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to identify published minimal clinically important difference (MID) thresholds relevant to this guideline. Identified MIDs were assessed to ensure they had been developed and validated in a methodologically rigorous way, and were applicable to the populations, interventions and outcomes specified in this guideline. In addition, the Guideline Committee were asked to prospectively specify any outcomes where they felt a consensus MID could be defined from their experience. MIDs identified through this process were intended to be used to inform discussions on the clinical importance of effects and the precision of effect estimates. No published MIDs were found through this process and the committee did not wish to pre specify consensus MIDs for any outcome. The clinical importance of effects was judged by the committee taking into account evidence across all outcomes and absolute effect estimates. These discussions are documented in the committee discussion section of each evidence review.

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.

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 5. Each element was graded using the levels of concern summarised in Table 6.

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 7. '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 all 4 components had 'no or very minor' concerns or 3 'no or very minor' and 1 'minor'. At the other extreme, a review finding would be downgraded 3 times (to 'very low') if at least 2 components had serious concerns or at least 3 had moderate concerns. A basic principle was that if any components had serious concerns then overall confidence in the review finding would be downgraded at least once (potentially more depending on the other ratings). Transparency about overall judgements is provided in the CERQual tables, including a brief reference to components for which there were concerns in the 'overall confidence' cell.

Table 5: Adaptation of GRADE quality elements for qualitative reviews

Quality element	Description
Risk of bias ('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.

Quality element	Description
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 6: 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 7: 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 8.

Table 8: 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 healthcare 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 rather to the depth of evidence and whether sufficient quotations or observations were provided to underpin the findings.

Reviewing economic evidence

Systematic reviews of economic literature were conducted for all review questions covered in the guideline.

Inclusion and exclusion of economic studies

Titles and abstracts of articles identified through the economic literature searches were assessed for inclusion using the predefined eligibility criteria listed in Table 9.

Table 9: Inclusion and exclusion criteria for systematic reviews of economic evaluations

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 economic information transferable to the UK context.
Intervention or comparators in accordance with the guideline scope
Study population in accordance with the guideline scope and review protocols for each review question
Full economic evaluations (cost-utility, cost effectiveness, cost-benefit or cost-consequence analyses) assessing both costs and outcomes associated with interventions of interest, as well as costing analyses that compared only costs between 2 or more interventions of interest were included in the review
Studies were included provided that sufficient details regarding methods and results were available to enable the methodological quality of the study to be assessed, and provided that the study's data and results were extractable.
Clinical effectiveness data utilised in the economic study should have been derived from a clinical trial, a prospective or retrospective cohort study (including before-after study designs), or from a literature review.
The outcome measure of the economic analysis should be the Quality Adjusted Life Year (QALY) or one of the measures considered in the clinical review.
Exclusion criteria
Poster presentations, conference abstracts and letters containing insufficient methodological details
Non-English language papers
Cost-of-illness type studies
Non-comparative studies
Studies that considered exclusively intervention costs, e.g. drug acquisition costs, without considering wider healthcare costs associated with the management of people self-harming

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 economic information transferable to the UK context.
Intervention or comparators in accordance with the guideline scope
Study population in accordance with the guideline scope
Full economic evaluations (cost-utility, cost effectiveness, cost-benefit or cost-consequence analyses) assessing both costs and outcomes associated with interventions of interest, as well as costing analyses that compared only costs between 2 or more interventions of interest were included in the review
Studies were included provided that sufficient details regarding methods and results were available to enable the methodological quality of the study to be assessed, and provided that the study's data and results were extractable.
Clinical effectiveness data utilised in the economic study should have been derived from a clinical trial, a prospective or retrospective cohort study, or from a literature review.
The outcome measure of the economic analysis should be the Quality Adjusted Life Year (QALY) or one of the measures considered in the clinical review.
Exclusion criteria
Poster presentations, conference abstracts and letters containing insufficient methodological details
Non-English language papers
Cost-of-illness type studies
Non-comparative studies
Studies that considered exclusively intervention costs, e.g. drug acquisition costs, without considering wider healthcare costs associated with the management of acne
Studies that compared costs of branded vs generic forms of the same drug

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

Eleven economic studies met inclusion criteria for the review. The PRISMA for the search of economic evaluations is presented in the appendix G of each evidence review. Summaries of economic evidence including economic evidence tables are presented in the respective evidence reports for each review question. Lists of economic studies excluded after obtaining full text with reasons for exclusion are provided in the appendix J of the relevant 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 (NICE 2020), 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; whereas studies rated as

either 'not applicable', with 'very serious limitations' or both were excluded from the committee discussion of the evidence.

Economic profiles of all economic studies that were considered during guideline development, including de novo economic analyses undertaken for this guideline, are provided in the heading 'Summary of included economic evidence' in the relevant evidence reviews.

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, as these need to be supported by robust evidence on cost effectiveness.

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 guideline committee prioritised the following review questions for economic modelling where it was thought that economic considerations would be particularly important in formulating recommendations:

- Cost-effectiveness of psychological and psychosocial interventions for people who have self-harmed. The methods and results of the 2 de novo economic analyses are fully reported in appendix I of evidence review J under the headings 'CBT-based psychotherapy for adults who have self-harmed' and 'DBT-A for children and young people who have self-harmed'.
- Cost-effectiveness associated with admission to acute general hospital for people who have self-harmed but no longer require physical care. This question was not possible to model due to lack of sufficient clinical evidence, as reported in evidence review T. For the same reason, this topic was later disregarded as a priority for bespoke economic modelling by the committee.

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

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

The committee's considerations of cost effectiveness are discussed explicitly under the heading '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 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.

References

Bradburn 2007

Bradburn, M. J., Deeks, J. J., Berlin, J. A., & Localio, A. R. Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events. Statistics in Medicine, 26, 53–77, 2007.

Dixon-Woods 2005

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

Hayden 2013

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

Higgins 2011

Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated 2019] The Cochrane Collaboration. Available from www.handbook.cochrane.org (accessed 10 November 2021)

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

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

NICE 2020

National Institute for Health and Care Excellence (NICE) (2014) NICE Policy on conflicts of interest (updated 2020). Available from https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-procedures/declaration-of-interests-policy.pdf (accessed 10 November 2021)

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

Witt 2021a

Witt KG, Hetrick SE, Rajaram G, Hazell P, Taylor Salisbury TL, Townsend E, Hawton K. Psychosocial interventions for self-harm in adults. Cochrane Database of Systematic Reviews 2021, Issue 4. Art. No.: CD013668. DOI: 10.1002/14651858.CD013668.pub2.

Witt 2021b

Witt KG, Hetrick SE, Rajaram G, Hazell P, Taylor Salisbury TL, Townsend E, Hawton K. Interventions for self-harm in children and adolescents. Cochrane Database of Systematic Reviews 2021, Issue 3. Art. No.: CD013667. DOI: 10.1002/14651858.CD013667.pub2.