National Institute for Health and Care Excellence

FINAL

Disabled children and young people up to 25 with severe complex needs: integrated service delivery and organisation across education, health and social care

Supplement A: Methods

NICE guideline NG213

Methods

March 2022

Final

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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

Contents

Development of the guideline

Remit

To see “What this guideline covers” and “What this guideline does not cover” please see the guideline scope Disabled children and young people up to 25 with severe complex needs.

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
- 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] Views and experiences of service users	What is the experience of disabled children and young people with severe complex needs and their families and carers of joint delivery of health, social care and education services?	Qualitative
[B] Involving children and young people	What are the most effective practices (for example, communication and information management) to enable health, social care and education services to work together to involve disabled children and young people with severe complex needs in understanding, planning and reviewing their care and education?	Intervention
[C] Combined approaches to identifying, assessing & monitoring needs	What are the most effective combined approaches to identifying, assessing and monitoring the health, social care and education needs (including changing needs) of disabled children and young people with severe complex needs	Intervention
[D] Supporting families and carers	What interventions, such as combined support, communication strategies and short breaks, are effective in enabling families and carers to be involved in the planning and delivery of care for disabled children and young people with severe complex needs?	Intervention
[E] Palliative and end of life care	What combined health, social care and education service delivery arrangements can best provide for the needs of disabled children and young people with severe complex	Intervention

Evidence review	Review question	Type of review
	needs on a palliative or advance care plan, and for the needs of their families and carers?	
[F] Supporting participation in education and social activities	What are the most effective ways that health, social care and education services can work together to support disabled children and young people with severe complex needs to participate in and benefit from education and social activities?	Intervention
[G] Promoting inclusion, independence and wellbeing	What are the most effective approaches for health, social care and education services to work together to promote inclusion, independence and wellbeing of disabled children and young people with severe complex needs?	Intervention
[H] Preparation for employment	What are the most effective models of health, social care and education services working together to prepare disabled children and young people with severe complex needs for employment?	Intervention
[I] Suitability and accessibility of environments	What are the most effective practices (for example, environmental assessments and use of equipment such as assistive technology across different contexts) to ensure the suitability and accessibility of the environments in which disabled children and young people with severe complex needs receive health and social care and education?	Intervention
[J] Planning and managing transition from children's to adults' services	What is the impact of including education with combined health and social care support models and frameworks on transition from children's to adults' services for disabled children and young people with severe complex needs?	Intervention
[K] Barriers and facilitators of joined-up care	What are the barriers and facilitators perceived or experienced by users and providers of joined-up care across health, social care, education and other services for disabled children and young people with severe complex needs?	Qualitative
[L] Enabling professionals to meet needs of children and young people	What are the most effective practices (for example, communication and training) to enable health, social care and education professionals to meet the combined health, social care and education needs of disabled children and young people with severe complex needs?	Intervention
[M] Views and experiences of service providers	What is the experience of commissioners and providers of joint working of health, social care and education services for disabled children and young people with severe complex needs?	Qualitative
[N] Commissioning, practice and service delivery models	• What are the most effective commissioning and practice models to deliver joined-up health, social care and education services for disabled children and young people with severe complex needs? • What combined service delivery models are most effective in meeting the health, social care and education needs (including changing and evolving needs) of disabled children and young people with severe complex needs?	Intervention

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 2 (Economics)
- Supplement 3 (NGA staff list).

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 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.

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: Medline, Medline-in-Process, Embase, Health Management Information Consortium (HMIC), Social Policy and Practice, PsycInfo, Emcare, Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessments (HTA), Applied Social Sciences Index and Abstracts (ASSIA), Social Services Abstracts, Sociological Abstracts, Educational Resources Information Centre (ERIC), British Education Index, Cumulative Index to Nursing and Allied Health (CINAHL plus), Social Sciences Citation Index (SSCI), and Social Care Online. All searches were restricted by date to 2000 onwards, as stated and explained in the individual review protocols for each review. The webpages of the following organisations were also checked for relevant publications for each review question: Kings Fund, National Audit Office, and Audit Commission.

Searches were run once for all reviews during development.

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 Health Technology Assessment (HTA) database. 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, Embase, Health Management Information Consortium (HMIC),

Social Policy and Practice, PsycInfo, Emcare, Cochrane Central Register of Controlled Trials (CCTR), Applied Social Sciences Index and Abstracts (ASSIA), Social Services Abstracts, Sociological Abstracts, Educational Resources Information Centre (ERIC), British Education Index, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Social Sciences Citation Index (SSCI) and Social Care Online. Where possible, searches were limited to studies published in English. All searches were restricted by date to 2000 onwards, as stated and explained in the individual review protocols for each review. The webpages of the following organisations were also checked for relevant economics publications: Kings Fund, National Audit Office and Audit Commission

As with the general literature searches, the economic literature searches were run once for all reviews during development.

Details of the search strategies, including the study-design filter used and databases searched, are provided in Supplement 2 (Health economics).

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

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.

Review questions, selected as high priorities for economic analysis (and those selected as medium priorities and where economic analysis could influence recommendations) and complex 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 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.

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. Where there was limited evidence from RCTs, non-randomised studies (NRS) were considered for inclusion. Service evaluations, process evaluations and audits were considered for inclusion in the absence of comparative non-randomised studies.

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 mortality, 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. Consequently, certain aspects of quality assessment such as imprecision of the effect estimate could not be assessed as per standard methods for this type of evidence and ratings based on sample size cut-offs were considered instead.

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).

When case series were included, descriptive data from the studies were included and no further analysis was performed.

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.

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.

Combining qualitative and quantitative evidence

The NGA technical team presented the data from quantitative and qualitative evidence reviews separately, however for most of the quantitative reviews there was also relevant qualitative evidence. The committee completed the synthesis of these mixed data through their discussions of the evidence, referring back to the qualitative evidence whenever considering the results of quantitative reviews. Where there was qualitative evidence that supported recommendations made for quantitative reviews, the supporting evidence is documented in the relevant quantitative review.

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 such as cross-sectional or before and after studies assessed using the Joanna Briggs Institute checklist for cross-sectional studies or the EPOC RoB tool, 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 version 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 version 2 of 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 the ROBIS checklist was used (see Appendix H in Developing NICE guidelines: the manual

For non-randomised 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.

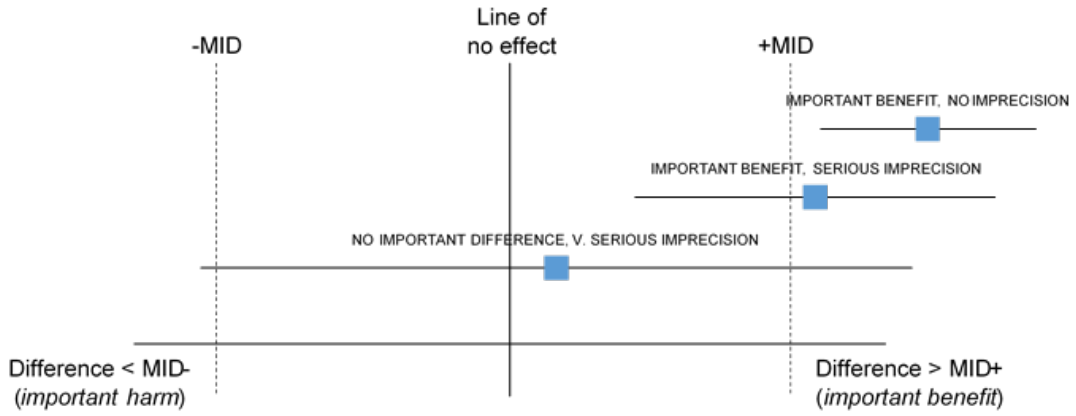
When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

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.

If risk difference was used for meta-analysis, for example if the majority of studies had zero events in either arm, imprecision was assessed based on sample size using 200 and 400 as cut-offs for very serious and serious imprecision respectively. These sample size cut-offs were also used to judge imprecision when results were presented as medians. The committee used these numbers based on commonly used optimal information size thresholds.

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

There were no meta-analyses of 3 or more studies in this guideline so funnel plots were not produced to assess potential for publication bias and evidence was not downgraded for publication bias.

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 2015) 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 7Table 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 or serious 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.

Transparency about overall judgements is provided in the CERQual tables, including a brief reference to components for which there were concerns in the 'level of concern' column.

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 confidence in review findings. 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 evidence 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
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. It is not equivalent to the number of studies contributing to a theme, but

Quality element	Description
	rather to the depth of evidence and whether sufficient quotations or observations were provided to underpin the findings.

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 10 areas summarised in Table 8.

Table 8: Methodological limitations in qualitative studies

Aims of the research	This domain assesses whether the aims, importance and relevance of the study were described clearly
Appropriateness of using qualitative methodology	This domain assesses whether qualitative research methods were appropriate for investigating the research question, for example, does the study aim to interpret or illuminate actions or subjective experiences
Research design	This domain assesses whether the study approach has been documented clearly and

	if it was justified, for example, based on a theoretical framework
Recruitment strategy	This domain assesses the procedure and reasons for the method of selecting participants and whether reasons for non-participation are discussed
Data collection	This domain assesses the documentation and justification of the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses where interviews took place, what form the data took (e.g., tape recordings, written notes) and data saturation
Relationship between researcher and participants	This domain assesses who conducted any interviews, any potential biases they might have and how these might have influenced the research questions or data collection. The assessment should include consideration of how the researcher responded to events during the study
Ethical considerations	This domain assesses whether ethical approval was obtained and ethical standards maintained, including issues of informed consent, confidentiality and the effect of the study on participants
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 whether contradictory data are taken into account and whether the researcher considered their own biases during analysis and selection of data for presentation also forms part of this assessment
Findings	This domain assesses whether findings are credible, reported explicitly and discussed in the context of the original research question. It also assesses if findings for and against the researchers' arguments are discussed
Value of research	This domain assesses if the researchers discuss the generalisability of findings, the contribution they make to existing knowledge and directions for future research

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 corresponds to the depth of evidence and whether sufficient quotations or observations were provided to underpin the findings. The complexity of the themes is also taken into account when assessing their adequacy. 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. Data would be considered thin where there is a lack of information and it is likely that further observations would provide more insight, and rich where there is sufficient information so further observations would be unlikely to suggest a different interpretation.

Reviewing economic evidence

A global economic literature search was undertaken to cover all review questions in the guideline.

Titles and abstracts of articles identified through the economic literature searches were independently 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
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. Cost analyses were also considered for inclusion due to the anticipated lack of economic evidence.
Only studies from Organisation for Economic Co-operation and Development countries were included.
Only studies published from 2000 onwards were included in the review.
Exclusion criteria
Abstracts containing insufficient methodological details.

Inclusion criteria
Cost-of-illness type studies.

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.

Details of economic evidence study selection, lists of excluded studies, economic evidence tables, the results of quality assessment of economic evidence (see below) and economic evidence profiles are presented in each of the evidence reports.

Appraising the quality of economic evidence

The quality of economic evidence was assessed using the economic evaluations checklist specified in Developing NICE guidelines: the manual. See the evidence reports for further details.

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 resources. Economic evaluations aim to integrate data on benefits 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 commissioners and so need special attention.

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.

- What are the most effective commissioning and practice models to deliver joined-up health, social care and education services for disabled children and young people with severe complex needs?
- What are the most effective combined approaches to identifying, assessing and monitoring the health, social care and education needs (including changing needs) of disabled children and young people with severe complex needs?
- What combined commissioning, practice and service delivery models are most effective in meeting the health, social care and education needs (including changing and evolving needs) of disabled children and young people with severe complex needs?
- What are the most effective approaches for health, social care and education services to work together to promote inclusion, independence and wellbeing of disabled children and young people with severe complex needs?

Original economic modelling was not undertaken for any review questions as there was insufficient effectiveness and cost data to inform new modelling. The committee was also of a view that care is very individual in this population and that any costings would not be generalizable.

Although, no modelling was undertaken the committee made a qualitative judgement regarding cost effectiveness by considering expected differences in resource and cost use between options, alongside effectiveness evidence.

Cost effectiveness criteria

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, however, it was acknowledged that this threshold may not be suitable for interventions that go beyond NHS and Personal Social Services (PSS) perspective
- 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 reports.

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 15 April 2021)

Lewin 2015

Lewin S, Glenton C, Munthe-Kaas H et al. (2015) Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS Med 12(10), e1001895

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 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 15 April 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