National Institute for Health and Care Excellence

Final

Rehabilitation of adults with complex psychosis

Supplement B: Methods

NICE guideline NG181

Development of the guideline and methods

August 2020

Final

This supplementary material was developed by the National Guideline Alliance which is part of 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-3828-5

Contents

Development of the guideline

Remit

The National Institute for Health and Care Excellence (NICE) commissioned the National Guideline Alliance (NGA) to develop a new guideline on rehabilitation in adults with complex psychosis and other severe mental health conditions.

For further details of what the guideline does and does not cover see: 'Guideline scope: Rehabilitation in adults with complex psychosis and related severe mental health conditions.'

The title of the guideline changed to "Rehabilitation for adults with complex psychosis" during development. The previous title of the guideline has been retained in the evidence reviews for consistency with the wording used in the review protocols.

Methods

Introduction

This section summarises methods used to identify and review the evidence, to consider cost effectiveness, and to develop guideline recommendations. This guideline was developed in accordance with methods described in the 2014 version of Developing NICE guidelines: the manual (NICE 2014).

Declarations of interest were recorded and managed in accordance with NICE's 2018 Policy on declaring and managing interests for NICE advisory committees.

Developing the review questions and outcomes

The review questions considered in 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:

- intervention reviews – using population, intervention, comparison and outcome (PICO)
- prognostic reviews – using population, presence or absence of a prognostic, risk or predictive factor and outcome (PPO)
- prevalence - using population and outcome (PO)
- qualitative reviews – using population, phenomenon of interest and context (PICo).

These frameworks guided the development of review protocols, the literature searching process, and critical appraisal and synthesis of evidence. They also facilitated development of recommendations by the committee.

Literature searches, critical appraisal and evidence reviews were completed for all review questions except for evidence review [I] Collaborative care planning.

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

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

Evidence review	Review question	Type of review
[A] Identifying people who would benefit most from mental health rehabilitation services	What service user characteristics are associated with successful progress in rehabilitation services for people with complex psychosis and related severe mental health conditions?	Prognostic

Evidence review	Review question	Type of review
[B] Barriers in accessing rehabilitation services	What coexisting medical, social (including family, cultural and ethnicity), communication, neurodevelopmental, cognitive or mental health problems pose barriers for people with complex psychosis and related severe mental health conditions in accessing rehabilitation services?	Qualitative
[C] Prevalence of comorbidity	What coexisting conditions (neurodevelopmental, cognitive, mental/physical health disorders) need to be considered when formulating a rehabilitation plan with people with complex psychosis?	Prevalence
[D] Effectiveness of rehabilitation services	What is the effectiveness of rehabilitation services compared with standard care?[1]	Intervention
[E] Comparative effectiveness of different types of rehabilitation services	What is the comparative effectiveness of different types of rehabilitation services?	Intervention
[F] Required components of an effective rehabilitation pathway	What are the required components of an effective rehabilitation pathway?	Prognostic
[G] Integrated rehabilitation care pathways involving multiple providers	What are the barriers and facilitators to integrated rehabilitation care pathways involving multiple providers (including health, social care, non-statutory, housing, independent and voluntary services)?	Qualitative
[H] Principles to guide adjustments to standard treatment	What principles should guide adjustments to standard treatments in the management of the underlying psychosis in people using rehabilitation services?	Intervention
[I] Collaborative care planning	What is the best way of involving people with complex psychosis and related severe mental health conditions, and their families and carers, in planning their care collaboratively with practitioners and providers?	Intervention
[J] The rehabilitation approaches, care, support and treatment that are valued by recipients	What rehabilitation approaches, care, support and treatment are valued by people with complex psychosis and related severe mental health conditions, and by their families and carers?	Qualitative
[K] Interventions to improve activities of daily living	What interventions specific to rehabilitation are effective for people with complex psychosis and other severe mental health conditions to improve their activities of daily living?	Intervention

Evidence review	Review question	Type of review
[L] Interventions to improve interpersonal functioning	What interventions specific to rehabilitation are effective for people with complex psychosis and related severe mental health conditions to improve their inter-personal functioning and social skills?	Intervention
[M] Interventions to improve engagement in community activities	What interventions specific to rehabilitation are effective for people with complex psychosis and related severe mental health conditions to improve their engagement in community activities (for example, leisure, education and work?[1]	Intervention
[N] Interventions to improve engagement in healthy living	What interventions specific to rehabilitation are effective in improving the engagement of people with complex psychosis and related severe mental health conditions in healthy living (nutrition weight, physical activity, sleep, oral health, accessing health services, health monitoring, smoking cessation)?	Intervention
[O] Effective interventions for improving engagement in addressing substance misuse	What interventions specific to rehabilitation are effective in improving the engagement of people with complex psychosis and other related severe mental health conditions in addressing substance misuse?	Intervention
[P] The features of supported accommodation and housing that promote successful community living	What features of supported accommodation and housing promote successful community living in people with complex psychosis and related severe mental health conditions?	Qualitative and Intervention
[Q] Factors associated with successful transition through rehabilitation services	What factors are associated with successful transition through rehabilitation services to other parts of the mental health, social care and primary care systems?	Prognostic
[R] Supporting successful transitions	What processes are needed to support successful transitions?	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 A (NGA staff list)
- Supplement B (Methods; this document)

1 Searching for evidence

2 Scoping search

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

5 Systematic literature search

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

8 Databases were searched using subject headings, free-text terms and, where
9 appropriate, study type filters. Where possible, searches were limited to retrieve
10 studies published in English. All the searches were conducted in the following
11 databases: Medline, Medline-in-Process, Embase, Psycinfo, Cochrane Central
12 Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews
13 (CDSR), Database of Abstracts of Reviews of Effects (DARE) and Health
14 Technology Assessments (HTA).

15 Searches were run once for all reviews during development except for the review
16 question about the best way of involving people with complex psychosis and related
17 severe mental health conditions, and their families and carers, in planning their care
18 collaboratively with practitioners and providers. The committee agreed that
19 collaborative care planning is a well-established requirement of all care planning in all
20 UK mental health services, with significant established guidance already in place.
21 Rather than review what was anticipated to be a sparse UK pool of evidence, the
22 committee agreed that their best approach would be to review the existing UK
23 guidance and adopt, adapt or refer to what is already in place.

24 Details of the search strategies, including the study-design filters used and
25 databases searched, are provided in appendix B of each evidence review. In
26 September 2019 the committee decided that none of the initial electronic searches
27 would need to be re-run and updated. All of the searches had been conducted less
28 than a year previously. For 11 of the searches [C, D, E, F, H, K, L, M, P (quantitative
29 search), Q and R] the committee agreed re-runs were unnecessary because they
30 had been searched less than 9 months previously. The qualitative search had been
31 conducted less than 1 year previously and had contributed between 9 and 21 studies
32 to each of the 4 qualitative reports [B, G, J, and P (qualitative)], and similarly 2
33 quantitative searches [A and N] were also conducted less than 1 year previously and
34 resulted in a significant amount of high quality studies. For this reason, the committee
35 were confident that re-runs would not raise any significant new data that would alter
36 the recommendations they had developed. One search [O] had been conducted 10
37 months previously and resulted in 1 randomised controlled trial (RCT), however upon
38 consulting their combined knowledge of recent research the committee felt confident
39 that no new trials had been published on this topic since then.

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

The economic literature searches were run once for all reviews during development.

Quality assurance

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

Reviewing 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 E of each evidence review).
- Included studies were critically appraised using an appropriate checklist as specified in the 2014 version of Developing NICE guidelines: the manual (NICE 2014). 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 checked 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. Studies with mixed populations were included if at least 66% of participants had complex psychosis or a related severe mental health condition. Given the importance of the cultural setting in which mental health rehabilitation takes place only studies from the UK, USA, Australasia, Europe and Canada were included because they have similar cultures to the UK.

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

For intervention reviews, 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 controlled trials were considered for inclusion.

For prognostic and prevalence reviews, prospective and retrospective cohort and case–control studies and case series were considered for inclusion. Studies that included multivariable analysis were prioritised.

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 D 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.

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 reviews

Pairwise meta-analysis

Meta-analysis to pool results from RCTs was conducted where possible using Cochrane Review Manager (RevMan5) software. Where non-randomised evidence was used, this was not meta-analysed.

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 duration of hospital stay, 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 subjective ratings were considered instead.

Subgroups for stratified analyses were agreed for some review questions as part of protocol development.

When meta-analysis was undertaken, the results were presented visually using forest plots generated using RevMan5 (see Appendix F 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 prognostic reviews

ORs or RRs with 95% CIs reported in published studies were extracted or calculated by the NGA technical team to examine relationships between risk factors and outcomes of interest. Ideally analyses would have adjusted for key confounders (such as age, or length of illness) to be considered for inclusion. If multiple results were deemed appropriate to meta-analyse (that is, there was sufficient similarity between risk factor and outcome under investigation) results were pooled using the generic inverse method. In most cases there was variation across studies in terms of populations, risk factors, outcomes and statistical analysis methods (including adjustments for confounding factors), and prognostic data were not pooled, but results from individual studies were presented in the evidence reviews.

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

Data synthesis for prevalence review

Prevalence rates with 95% CIs reported in published studies were extracted. For each condition of interest the most recent and relevant population based study was used to provide the prevalence estimate, and no further synthesis was done.

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.

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 start as 'high' quality evidence and non-randomised studies 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

Quality element	Description
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 (see Appendix H in Developing NICE guidelines: the manual; NICE 2014).

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

- selection bias
- performance bias
- attrition bias

- detection bias
- reporting bias.

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 Newcastle-Ottawa 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 considerable heterogeneity, and more than 80% indicating very serious heterogeneity. When considerable 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 considerable heterogeneity was present, 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.

When no plausible explanation for the heterogeneity could be found, the quality of the evidence was downgraded in GRADE for inconsistency.

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 2 treatments. Three decision-making zones can be differentiated, bounded by the thresholds for minimal importance (minimally important differences; MIDs) for benefit and harm.

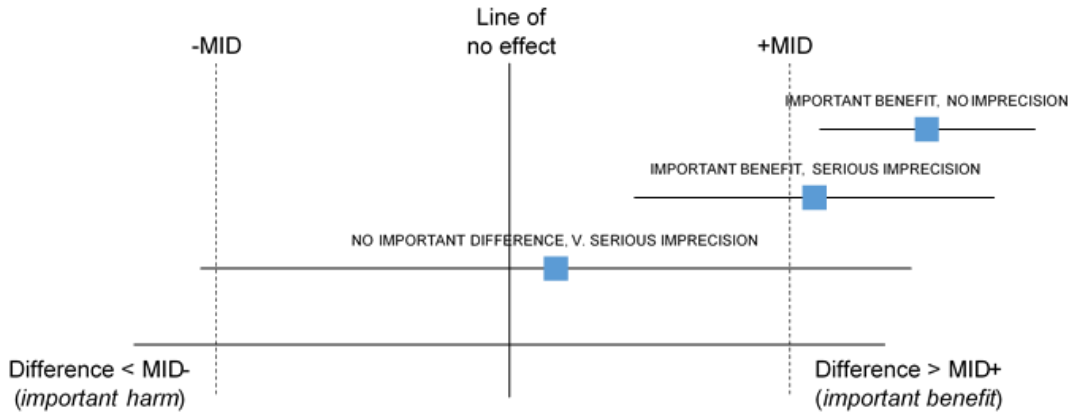
When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

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

Assessing publication bias in intervention reviews

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

Prognostic studies

Adapted GRADE methodology for prognostic reviews

For prognostic reviews with evidence from comparative studies an adapted GRADE approach was used. As noted above, GRADE methodology is designed for intervention reviews but the quality assessment elements were adapted for prognostic reviews.

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

Table 5: Adaptation of GRADE quality elements for prognostic reviews

Quality element	Description
Risk of bias ('Study limitations')	Limitations in study design and implementation may bias estimates and interpretation of the effect of the prognostic/risk factor. High risk of bias for the majority of the evidence reduces confidence in the estimated effect. Prognostic studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)
Inconsistency	This refers to unexplained heterogeneity between studies looking at the same prognostic/risk factor, resulting in wide variability in estimates of association (such as RRs or ORs), with little or no overlap in confidence intervals
Indirectness	This refers to any departure from inclusion criteria listed in the review protocol (such as differences in study populations or prognostic/risk factors), that may affect the generalisability of results
Imprecision	This occurs when a study has relatively few participants and also when the number of participants is too small for a multivariable analysis (as a rule of thumb, 10 participants are needed per variable). This was assessed by considering the confidence interval in relation to the point estimate for each outcome reported in the included studies

RR, relative risk; OR, odds ratio

Assessing risk of bias in prognostic reviews

The Quality in Prognosis Studies (QUIPS) tool developed by Hayden 2013 was used to assess risk of bias in studies included in prognostic reviews (see Appendix H in Developing NICE guidelines: the manual; NICE 2014). The risk of bias in each study was determined by assessing the following domains:

- selection bias
- attrition bias
- prognostic factor bias
- outcome measurement bias
- control for confounders
- appropriate statistical analysis.

Assessing inconsistency in prognostic reviews

If multiple results were deemed appropriate to meta-analyse (that is, there was sufficient similarity between risk factor and outcome under investigation) inconsistency was assessed by visually inspecting forest plots and observing whether there was considerable heterogeneity in the results of the meta-analysis. This was assessed by calculating the I-squared statistic for the meta-analysis with an I-squared value of more than 50% indicating considerable heterogeneity, and more than 80% indicating very serious heterogeneity. When considerable or very serious heterogeneity was observed, possible reasons were explored and subgroup analyses were performed as pre-specified in the review protocol where possible.

When no plausible explanation for the heterogeneity could be found, the quality of the evidence was downgraded in GRADE for inconsistency.

Assessing indirectness in prognostic reviews

Indirectness in prognostic reviews was assessed by comparing the populations, prognostic factors and outcomes in the evidence to those defined in the review protocol.

Assessing imprecision and importance in prognostic reviews

Prognostic studies may have a variety of purposes, for example, establishing typical prognosis in a broad population, establishing the effect of patient characteristics on prognosis, and developing a prognostic model. While by convention MIDs relate to intervention effects, the committee agreed to use GRADE default MIDs for intervention studies as a starting point from which to assess whether the size of an outcome effect in a prognostic study would be large enough to be meaningful in practice. For effect measures without a GRADE default MID (such as odds ratios) a sample size criterion was used (as a rule of thumb, 10 participants are needed per variable in a multivariable prognostic model) or the quality was downgraded one level because imprecision could not be assessed.

Prevalence review

GRADE methods were not used for the prevalence review. The quality of evidence about the prevalence of each condition was based on the critical appraisal of the individual studies.

Qualitative reviews

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 6. Each element was graded using the levels of concern summarised in Table 7. The ratings for each component were combined (as with other types of evidence) to obtain an overall assessment of quality for each theme as described in Table 8.

Table 6: 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. Individual studies that may have contributed to a theme or sub-theme may have been conducted in a manner that by design would have not reached theoretical saturation at an individual study level

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

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

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

Overall confidence level	Definition
High	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest
Moderate	It is likely that the review finding is a reasonable representation of the phenomenon of interest

Overall confidence level	Definition
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; NICE 2014). Overall methodological limitations were derived by assessing the methodological limitations across the 6 domains summarised in Table 9.

Table 9: Methodological limitations in qualitative studies

Aim and appropriateness of qualitative evidence	This domain assesses whether the aims and relevance of the study were described clearly and whether qualitative research methods were appropriate for investigating the research question
Rigour in study design or validity of theoretical approach	This domain assesses whether the study approach was documented clearly and whether it was based on a theoretical framework (such as ethnography or grounded theory). This does not necessarily mean that the framework has to be stated explicitly, but a detailed description ensuring transparency and reproducibility should be provided
Sample selection	This domain assesses the background, the procedure and reasons for the method of selecting participants. The assessment should include consideration of any relationship between the researcher and the participants, and how this might have influenced the findings
Data collection	This domain assesses the documentation of the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses who conducted any interviews, how long they lasted and where they took place
Data analysis	This domain assesses whether sufficient detail was documented for the analytical

	process and whether it was in accordance with the theoretical approach. For example, if a thematic analysis was used, the assessment would focus on the description of the approach used to generate themes. Consideration of data saturation would also form part of this assessment (it could be reported directly or it might be inferred from the citations documented that more themes could be found)
Results	This domain assesses any reasoning accompanying reporting of results (for example, whether a theoretical proposal or framework is provided)

Assessing relevance of evidence in qualitative reviews

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

Assessing coherence of findings in qualitative reviews

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

Assessing importance in qualitative reviews

For themes stemming from qualitative findings, importance was agreed by the committee taking account of the generalisability of the context from which the theme was derived and whether it was sufficiently convincing to support or warrant a change in current practice, as well as the quality of the evidence.

Reviewing economic evidence

Inclusion and exclusion of economic studies

Titles and abstracts of articles identified through the economic literature searches were independently assessed for inclusion using the predefined eligibility criteria listed in Table 10.

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

Inclusion criteria
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
Exclusion criteria
Abstracts containing insufficient methodological details
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 health 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 (NICE 2014).

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 with the costs of different options. In addition, the

economic input aimed to identify areas of high resource impact; these are recommendations which (while cost effective) might have a large impact on Clinical Commissioning Group or Trust finances and so need special attention.

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 is the effectiveness of rehabilitation services compared with standard care?
- What is the comparative effectiveness of different types of rehabilitation services?
- What interventions specific to rehabilitation are effective for people with complex psychosis and related severe mental health conditions to improve their engagement in community activities (for example, leisure, education and work)?

A simple cost utility analysis was undertaken for the review question about the effectiveness of rehabilitation services as there was insufficient effectiveness data obtained from the clinical review that allowed for all relevant costs and outcomes to be considered.

A cost analysis was conducted for the review question about the comparative effectiveness of different types of rehabilitation services as there was insufficient evidence on the outcome 'out-of-area placements' to allow modelling a cost utility analysis. Evidence obtained from the grey literature provided most of the model inputs. A full cost utility analysis was undertaken for the review question regarding which interventions specific to rehabilitation services are effective for people with complex psychosis. This was an intervention type question, which lend themselves well to economic modelling, and effectiveness data was obtained from the accompanying clinical review. The methods and results of the de novo economic analyses are reported in Appendix J of the relevant evidence reports. When 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 Social value judgements: principles for the development of NICE guidance 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.

Details of the cost effectiveness analyses undertaken for the guideline are presented in Appendix J of the evidence reports.

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 (NICE 2014).

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 (NICE 2014).

Validation process

This guideline will be subject to a 6-week public consultation and feedback process. All comments received from registered stakeholders will be responded to in writing and posted on the NICE website at publication. For further details refer to Developing NICE guidelines: the manual (NICE 2014).

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 (NICE 2014).

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 March 2011] The Cochrane Collaboration. Available from www.handbook.cochrane.org (accessed 15 June 2018)

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 2014

National Institute for Health and Care Excellence (NICE) (2014) Developing NICE guidelines: the manual. Available from https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf (accessed 1 November 2019)

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 1 November 2019)

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